Summary

1. loc :- Label wise iloc :- Index wise

2. Value_counts() methods : It works on always series type of data

3. The Pandas Plot() : .plot(kind="bar")

4. sort_values() : It apply on both series and datafram

5. Drop Duplicate, It gives only unique value of cities : df.drop_duplicates(subset="city")

6. Groupby , Analysis over group cities = df.groupby("city"): cities.get_group("Hyderabad")

7. Merge two dataset: df.merge(df1,left_on="id",right_on="t_id")

8. Pivot Table: df.pivot_table(index="season",columns="winner",values="win_by_runs",aggfunc="count")

9. Set_index and reset_index: df.set_index("id")

10. Handle Missing Values

(i). Drop missing values using dropna() df.dropna(subset=["umpire1"]): Find any row with NAN values remove that whole row

(ii). Filling Missing Values With fillna(): df["season"].fillna(df["season"].mean())

In [469]:

```python
import numpy as np
import pandas as pd
```

In [3]:

```python
df  = pd.read_csv("matches - matches.csv")
```

In [7]:

```python
# head() is a function to get data priview 0:-1
df.head()
```

Out[7]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	

In [6]:

```
# In pandas tabular data is called DataFrame
type(df)
```

Out[6]:

```
pandas.core.frame.DataFrame
```

In [10]:

```
# shape is an attribute
df.shape
```

Out[10]:

```
(636, 18)
```

In [12]:

```
# In pandas object == string
df.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 [23]:

```
# It works for only numerical columns
df.describe()
```

Out[23]:

	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	0.156604	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 [31]:

```
# To fetch multiple column so we pass column name inside list
df[["team1","team2","winner"]]
```

Out[31]:

	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 [32]:

```
df[["team1","team2","winner"]].shape
```

Out[32]:

(636, 3)

In [36]:

```
# Datatype would be series because we fetch only one row
df.iloc[0]
```

Out[36]:

```
id                                                    1
season                                             2017
city                                          Hyderabad
date                                         2017-04-05
team1                               Sunrisers Hyderabad
team2                       Royal Challengers Bangalore
toss_winner                 Royal Challengers Bangalore
toss_decision                                     field
result                                           normal
dl_applied                                            0
winner                              Sunrisers Hyderabad
win_by_runs                                          35
win_by_wickets                                        0
player_of_match                            Yuvraj Singh
venue             Rajiv Gandhi International Stadium, Uppal
umpire1                                      AY Dandekar
```

```
umpire2                                        NJ Llong
umpire3                                             NaN
Name: 0, dtype: object
```

In [38]:

```python
# Datatype would be DataFrame because we fetch multiple rows
df.iloc[0:2]
```

Out[38]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	

In [39]:

```python
df.iloc[1:11:4]
```

Out[39]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	
5	6	2017	Hyderabad	2017-04-09	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	
9	10	2017	Mumbai	2017-04-12	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	

In [41]:

```python
# Fancy Indexing
df.iloc[[1,5,9]]
```

Out[41]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	
5	6	2017	Hyderabad	2017-04-09	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	
9	10	2017	Mumbai	2017-04-12	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	

In [45]:

```python
df.iloc[:,[4,5,10 ]]
```

Out[45]:

	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 [ ]:

In [55]:

```
# How many matches is held into Hyderabad city
mask = df["city"]=="Hyderabad"
```

In [58]:

```
df[mask].shape[0]
```

Out[58]:

```
49
```

In [59]:

```
# Creating Function
def get_match_city(city):
    mask = df["city"]==city
    return df[mask].shape[0]
```

In [61]:

```
get_match_city("Bangalore")
```

Out[61]:

```
66
```

In [62]:

```
get_match_city("Mumbai")
```

Out[62]:

```
85
```

In [73]:

```
# After 2010 how many match held in Hydrabad city.
mask1 =df["date"]>"2010-01:01"
mask2 =df["city"]=="Hyderabad"
df[mask1 & mask2].shape[0]
```

Out[73]:

42

```python
# Creating Function
def get_city_month(date,city):
    mask1 =df["date"]>date
    mask2 =df["city"]==city
    return df[mask1 & mask2].shape[0]
```

In [80]:

```python
get_city_month("2010-01:01","Hyderabad")
```

Out[80]:

42

Value_counts() methods

It works on always series type of data

In [81]:

```python
df.head(1)
```

Out[81]:

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

In [84]:

```python
# Value_counts method count the value of the particular column
df["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
```

In [86]:

```python
df["city"].value_counts()
```

Out[86]:

```
Mumbai       85
Bangalore    66
Kolkata      61
Delhi        60
Hyderabad    49
```

```
Hyderabad              49
Chennai               48
Chandigarh            46
Jaipur                33
Pune                  32
Durban                15
Ahmedabad             12
Centurion             12
Visakhapatnam         11
Rajkot                10
Dharamsala             9
Johannesburg           8
Port Elizabeth         7
Abu Dhabi              7
Cape Town              7
Cuttack                7
Ranchi                 7
Sharjah                6
Raipur                 6
Kochi                  5
Indore                 5
Kanpur                 4
East London            3
Nagpur                 3
Kimberley              3
Bloemfontein           2
Name: city, dtype: int64
```

In [87]:

```python
df["win_by_runs"].value_counts()
```

Out[87]:

```
0       349
7         9
23        9
10        9
9         9
       ...
60        1
59        1
50        1
57        1
146       1
Name: win_by_runs, Length: 86, dtype: int64
```
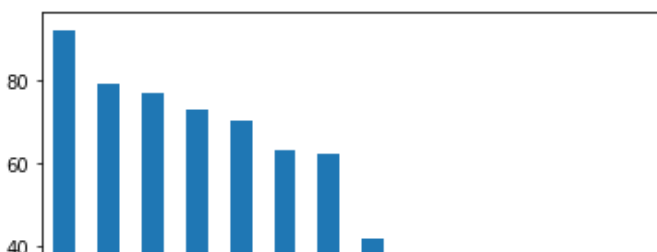
The Pandas Plot()

In [88]:

```python
# If you don't import this library you graph will be created
# but not display in the jupyter notebook.
import matplotlib.pyplot as plt
```
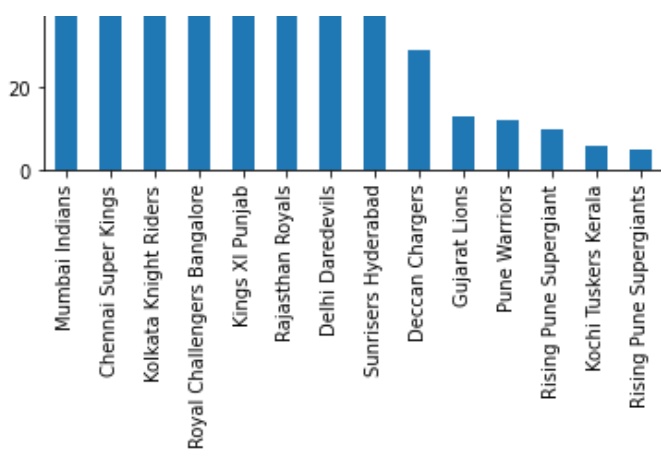
In [109]:

```python
df["winner"].value_counts().plot(kind="bar") #bar is used in categorical
```
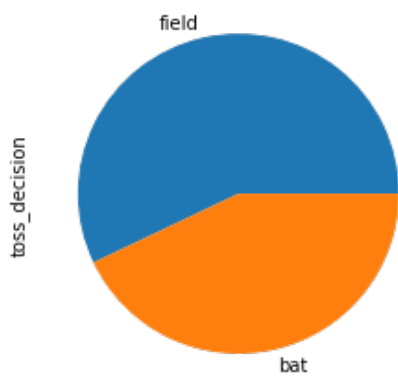
Out[109]:

```
<AxesSubplot:>
```

In [115]:

```python
df["toss_decision"].value_counts().plot(kind="pie")
```

Out[115]:

```
<AxesSubplot:ylabel='toss_decision'>
```


In [117]:
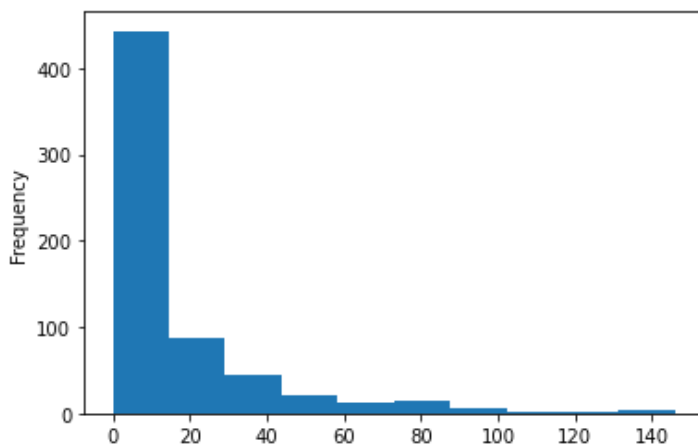
```python
df["win_by_runs"].plot(kind="hist")
```

Out[117]:

```
<AxesSubplot:ylabel='Frequency'>
```


In [ ]:

SERIES

In [119]:

```python
winner = df["winner"].value_counts()
```

In [123]:

```
winner
```

Out[123]:

```
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
```

In [124]:

```
winner.index
```

Out[124]:

```
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 [125]:

```
winner.values
```

Out[125]:

```
array([92, 79, 77, 73, 70, 63, 62, 42, 29, 13, 12, 10,  6,  5])
```

In [128]:

```
# How many matches play by all team1 and team2
# If index are same then we simply add them
df["team1"].value_counts()  + df["team2"].value_counts()
```

Out[128]:

```
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
```

sort_values()

It apply on both series and datafram

It apply on both series and datafram

In [131]:

```python
(df["team1"].value_counts()  + df["team2"].value_counts()).sort_values(ascending=False)
```

Out[131]:

```
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 [132]:

```python
team = df["team1"].value_counts()  + df["team2"].value_counts()
```

In [137]:

```python
team.sort_values(ascending=False)
```

Out[137]:

```
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 [143]:

```python
df.sort_values("city")
# If you want changes is happen in original dataset then you write
# inplace called True
# df.sort_values("city", inplace=True)
```

Out[143]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_
460	461	2014	Abu Dhabi	2014-04-18	Sunrisers Hyderabad	Rajasthan Royals	Rajasthan Royals	field	normal	0	Rajasthan Royals	
471	472	2014	Abu Dhabi	2014-04-26	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kings XI Punjab	
470	471	2014	Abu Dhabi	2014-04-26	Royal Challengers Bangalore	Rajasthan Royals	Rajasthan Royals	field	normal	0	Rajasthan Royals	
			Abu	2014-	Chennai	Delhi	Chennai				Chennai	

	464	465	2014	Dhabi	04-21	Super Kings	Daredevils	Super Kings	bat	normal	0	Super Kings
	id	season	city	date	team1	team2	toss_winner	toss_decision	normal result	dl_applied	winner	win_by_r
459	460	2014	Abu Dhabi	2014-04-18	Chennai Super Kings	Kings XI Punjab	Chennai Super Kings	bat	normal	0	Kings XI Punjab	
...	...	...	...	...	...	...	...	...	...	...	...	
466	467	2014	NaN	2014-04-23	Chennai Super Kings	Rajasthan Royals	Rajasthan Royals	field	normal	0	Chennai Super Kings	
468	469	2014	NaN	2014-04-25	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad	
469	470	2014	NaN	2014-04-25	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	0	Chennai Super Kings	
474	475	2014	NaN	2014-04-28	Royal Challengers Bangalore	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	
476	477	2014	NaN	2014-04-30	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	0	Sunrisers Hyderabad	

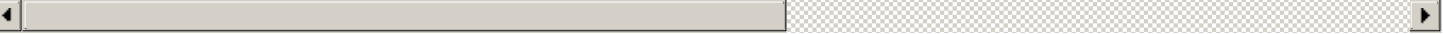**636 rows × 18 columns**

In [152]:

```
# Sort data according to date and city.
df.sort_values(["city","date"])
```

Out[152]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_r
457	458	2014	Abu Dhabi	2014-04-16	Kolkata Knight Riders	Mumbai Indians	Kolkata Knight Riders	bat	normal	0	Kolkata Knight Riders	
459	460	2014	Abu Dhabi	2014-04-18	Chennai Super Kings	Kings XI Punjab	Chennai Super Kings	bat	normal	0	Kings XI Punjab	
460	461	2014	Abu Dhabi	2014-04-18	Sunrisers Hyderabad	Rajasthan Royals	Rajasthan Royals	field	normal	0	Rajasthan Royals	
464	465	2014	Abu Dhabi	2014-04-21	Chennai Super Kings	Delhi Daredevils	Chennai Super Kings	bat	normal	0	Chennai Super Kings	
470	471	2014	Abu Dhabi	2014-04-26	Royal Challengers Bangalore	Rajasthan Royals	Rajasthan Royals	field	normal	0	Rajasthan Royals	
...	...	...	...	...	...	...	...	...	...	...	...	
466	467	2014	NaN	2014-04-23	Chennai Super Kings	Rajasthan Royals	Rajasthan Royals	field	normal	0	Chennai Super Kings	
468	469	2014	NaN	2014-04-25	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad	
469	470	2014	NaN	2014-	Mumbai	Chennai Super	Mumbai	bat	normal	0	Chennai Super	

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_r
474	475	2014	NaN	2014-04-28	Royal Challengers Bangalore	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	
476	477	2014	NaN	2014-04-30	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	0	Sunrisers Hyderabad	

636 rows × 18 columns

```
In [153]:
df.sort_values(["city","date"],ascending=[False,True])
Out[153]:
```

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner
312	313	2012	Visakhapatnam	2012-04-07	Chennai Super Kings	Deccan Chargers	Deccan Chargers	field	normal	0	Chennai Super Kings
315	316	2012	Visakhapatnam	2012-04-09	Deccan Chargers	Mumbai Indians	Deccan Chargers	bat	normal	0	Mumbai Indians
528	529	2015	Visakhapatnam	2015-04-16	Sunrisers Hyderabad	Rajasthan Royals	Rajasthan Royals	field	normal	0	Rajasthan Royals
530	531	2015	Visakhapatnam	2015-04-18	Delhi Daredevils	Sunrisers Hyderabad	Delhi Daredevils	bat	normal	0	Delhi Daredevils
536	537	2015	Visakhapatnam	2015-04-22	Sunrisers Hyderabad	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	1	Sunrisers Hyderabad
...	...	...	...	...	...	...	...	...	...	...	...
466	467	2014	NaN	2014-04-23	Chennai Super Kings	Rajasthan Royals	Rajasthan Royals	field	normal	0	Chennai Super Kings
468	469	2014	NaN	2014-04-25	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad
469	470	2014	NaN	2014-04-25	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat	normal	0	Chennai Super Kings
474	475	2014	NaN	2014-04-28	Royal Challengers Bangalore	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab
476	477	2014	NaN	2014-04-30	Sunrisers Hyderabad	Mumbai Indians	Mumbai Indians	field	normal	0	Sunrisers Hyderabad

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner

636 rows × 18 columns

In [ ]:

Drop Duplicate

In [155]:

```
df.drop_duplicates(subset="city")
```

Out[155]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winne
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrise Hyderaba
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pur Supergia
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolka Knig Ride
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings Punja
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Roy Challenge Bangalo
6	7	2017	Mumbai	2017-04-09	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal	0	Mumb Indiar
10	11	2017	Kolkata	2017-04-13	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolka Knig Ride
14	15	2017	Delhi	2017-04-15	Delhi Daredevils	Kings XI Punjab	Delhi Daredevils	bat	normal	0	Del Daredevi
31	32	2017	Chandigarh	2017-04-28	Sunrisers Hyderabad	Kings XI Punjab	Kings XI Punjab	field	normal	0	Sunrise Hyderaba
48	49	2017	Kanpur	2017-05-10	Gujarat Lions	Delhi Daredevils	Delhi Daredevils	field	normal	0	Del Daredevi
64	65	2008	Jaipur	2008-04-21	Kings XI Punjab	Rajasthan Royals	Kings XI Punjab	bat	normal	0	Rajasthan Roya
66	67	2008	Chennai	2008-04-23	Chennai Super Kings	Mumbai Indians	Mumbai Indians	field	normal	0	Chenn Sup King
117	118	2009	Cape Town	2009-04-18	Mumbai Indians	Chennai Super Kings	Chennai Super Kings	field	normal	0	Mumb Indiar
121	122	2009	Port Elizabeth	2009-04-20	Chennai Super	Royal Challengers	Chennai Super	bat	normal	0	Chenn Sup

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner
122	123	2009	Durban	2009-04-21	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	1	Kolkata Knight Riders
132	133	2009	Centurion	2009-04-28	Delhi Daredevils	Rajasthan Royals	Delhi Daredevils	bat	normal	0	Rajasthan Royals
137	138	2009	East London	2009-05-01	Mumbai Indians	Kolkata Knight Riders	Mumbai Indians	bat	normal	0	Mumbai Indians
140	141	2009	Johannesburg	2009-05-02	Chennai Super Kings	Delhi Daredevils	Delhi Daredevils	field	normal	0	Chennai Super Kings
150	151	2009	Kimberley	2009-05-09	Deccan Chargers	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab
160	161	2009	Bloemfontein	2009-05-15	Delhi Daredevils	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab
179	180	2010	Ahmedabad	2010-03-15	Rajasthan Royals	Delhi Daredevils	Delhi Daredevils	field	normal	0	Delhi Daredevils
185	186	2010	Cuttack	2010-03-19	Deccan Chargers	Kings XI Punjab	Kings XI Punjab	field	normal	0	Deccan Chargers
209	210	2010	Nagpur	2010-04-05	Rajasthan Royals	Deccan Chargers	Rajasthan Royals	bat	normal	0	Rajasthan Royals
224	225	2010	Dharamsala	2010-04-16	Kings XI Punjab	Deccan Chargers	Deccan Chargers	field	normal	0	Deccan Chargers
236	237	2011	Kochi	2011-04-09	Kochi Tuskers Kerala	Royal Challengers Bangalore	Kochi Tuskers Kerala	bat	normal	0	Royal Challengers Bangalore
312	313	2012	Visakhapatnam	2012-04-07	Chennai Super Kings	Deccan Chargers	Deccan Chargers	field	normal	0	Chennai Super Kings
418	419	2013	Raipur	2013-04-28	Delhi Daredevils	Pune Warriors	Pune Warriors	field	normal	0	Delhi Daredevils
440	441	2013	Ranchi	2013-05-12	Royal Challengers Bangalore	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders
457	458	2014	Abu Dhabi	2014-04-16	Kolkata Knight Riders	Mumbai Indians	Kolkata Knight Riders	bat	normal	0	Kolkata Knight Riders
458	459	2014	Sharjah	2014-04-17	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Royal Challengers Bangalore
461	462	2014	NaN	2014-04-19	Mumbai Indians	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Royal Challengers Bangalore

In [159]:

```python
# All duplicate values will be drop
df.drop_duplicates(subset="city").shape
```

(31, 18)

```python
df.drop_duplicates(subset=["season","city"]).shape
```

(107, 18)

```python
df.drop_duplicates(subset="season",keep="last")[["season","winner"]].sort_values("season")
```

	season	winner
116	2008	Rajasthan Royals
173	2009	Deccan Chargers
233	2010	Chennai Super Kings
306	2011	Chennai Super Kings
380	2012	Kolkata Knight Riders
456	2013	Mumbai Indians
516	2014	Kolkata Knight Riders
575	2015	Mumbai Indians
635	2016	Sunrisers Hyderabad
58	2017	Mumbai Indians

```python
df.head()
```

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

```python
(df.drop_duplicates(subset="season",keep="last")[["season","winner"]]).sort_values("seaso
```

```
n")
```

Out[180]:

	season	winner
116	**2008**	**Rajasthan Royals**
173	**2009**	**Deccan Chargers**
233	**2010**	**Chennai Super Kings**
306	**2011**	**Chennai Super Kings**
380	**2012**	**Kolkata Knight Riders**
456	**2013**	**Mumbai Indians**
516	**2014**	**Kolkata Knight Riders**
575	**2015**	**Mumbai Indians**
635	**2016**	**Sunrisers Hyderabad**
58	**2017**	**Mumbai Indians**

In [182]:

```
df.head(1)
```

Out[182]:

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

Groupby

Analysis over group

In [184]:

```
cities = df.groupby("city")
```

In [185]:

```
cities
```

Out[185]:

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f82a15b4190>
```

In [190]:

```
type(cities)
```

Out[190]:

```
pandas.core.groupby.generic.DataFrameGroupBy
```

In [192]:

```
# Number or rows
len(df)
```

Out[192]:

```
636
```

```
In [194]:
```

```
len(cities)
```

```
Out[194]:
```

```
30
```

```
In [197]:
```

```
cities.size().sort_values(ascending=False)
```

```
Out[197]:
```

```
city
Mumbai            85
Bangalore         66
Kolkata           61
Delhi             60
Hyderabad         49
Chennai           48
Chandigarh        46
Jaipur            33
Pune              32
Durban            15
Centurion         12
Ahmedabad         12
Visakhapatnam     11
Rajkot            10
Dharamsala         9
Johannesburg       8
Cape Town          7
Cuttack            7
Ranchi             7
Port Elizabeth     7
Abu Dhabi          7
Raipur             6
Sharjah            6
Indore             5
Kochi              5
Kanpur             4
Nagpur             3
Kimberley          3
East London        3
Bloemfontein       2
dtype: int64
```

```
In [458]:
```

```
# In winner column also city name so count it and tell the total number
cities["winner"].count().sort_values(ascending=False)
```

```
Out[458]:
```

```
city
Mumbai            85
Bangalore         64
Kolkata           61
Delhi             59
Hyderabad         49
Chennai           48
Chandigarh        46
Jaipur            33
Pune              32
Durban            15
Centurion         12
Ahmedabad         12
Visakhapatnam     11
Rajkot            10
Dharamsala         9
Johannesburg       8
Cape Town          7
Cuttack            7
```

```
Ranchi                  7
Port Elizabeth          7
Abu Dhabi               7
Raipur                  6
Sharjah                 6
Indore                  5
Kochi                   5
Kanpur                  4
Nagpur                  3
Kimberley               3
East London             3
Bloemfontein            2
Name: winner, dtype: int64
```

```
df["winner"].head()
```

```
0            Sunrisers Hyderabad
1          Rising Pune Supergiant
2          Kolkata Knight Riders
3                 Kings XI Punjab
4    Royal Challengers Bangalore
Name: winner, dtype: object
```

```
#df["city"].value_counts()
```

```
cities.first()
```

city	id	season	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	w
Abu Dhabi	458	2014	2014-04-16	Kolkata Knight Riders	Mumbai Indians	Kolkata Knight Riders	bat	normal	0	Kolkata Knight Riders	
Ahmedabad	180	2010	2010-03-15	Rajasthan Royals	Delhi Daredevils	Delhi Daredevils	field	normal	0	Delhi Daredevils	
Bangalore	5	2017	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	
Bloemfontein	161	2009	2009-05-15	Delhi Daredevils	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	
Cape Town	118	2009	2009-04-18	Mumbai Indians	Chennai Super Kings	Chennai Super Kings	field	normal	0	Mumbai Indians	
Centurion	133	2009	2009-04-28	Delhi Daredevils	Rajasthan Royals	Delhi Daredevils	bat	normal	0	Rajasthan Royals	
Chandigarh	32	2017	2017-04-28	Sunrisers Hyderabad	Kings XI Punjab	Kings XI Punjab	field	normal	0	Sunrisers Hyderabad	
Chennai	67	2008	2008-04-23	Chennai Super Kings	Mumbai Indians	Mumbai Indians	field	normal	0	Chennai Super Kings	
Cuttack	186	2010	2010-03-19	Deccan Chargers	Kings XI Punjab	Kings XI Punjab	field	normal	0	Deccan Chargers	

city	id	season	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	w
Dharamsala	225	2010	2010-04-16	Kings XI Punjab	Deccan Chargers	Deccan Chargers	field	normal	0	Deccan Chargers	
Durban	123	2009	2009-04-21	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	1	Kolkata Knight Riders	
East London	138	2009	2009-05-01	Mumbai Indians	Kolkata Knight Riders	Mumbai Indians	bat	normal	0	Mumbai Indians	
Hyderabad	1	2017	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	
Indore	4	2017	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	
Jaipur	65	2008	2008-04-21	Kings XI Punjab	Rajasthan Royals	Kings XI Punjab	bat	normal	0	Rajasthan Royals	
Johannesburg	141	2009	2009-05-02	Chennai Super Kings	Delhi Daredevils	Delhi Daredevils	field	normal	0	Chennai Super Kings	
Kanpur	49	2017	2017-05-10	Gujarat Lions	Delhi Daredevils	Delhi Daredevils	field	normal	0	Delhi Daredevils	
Kimberley	151	2009	2009-05-09	Deccan Chargers	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	
Kochi	237	2011	2011-04-09	Kochi Tuskers Kerala	Royal Challengers Bangalore	Kochi Tuskers Kerala	bat	normal	0	Royal Challengers Bangalore	
Kolkata	11	2017	2017-04-13	Kings XI Punjab	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	
Mumbai	7	2017	2017-04-09	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	
Nagpur	210	2010	2010-04-05	Rajasthan Royals	Deccan Chargers	Rajasthan Royals	bat	normal	0	Rajasthan Royals	
Port Elizabeth	122	2009	2009-04-20	Chennai Super Kings	Royal Challengers Bangalore	Chennai Super Kings	bat	normal	0	Chennai Super Kings	
Pune	2	2017	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	
Raipur	419	2013	2013-04-28	Delhi Daredevils	Pune Warriors	Pune Warriors	field	normal	0	Delhi Daredevils	
Rajkot	3	2017	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	
Ranchi	441	2013	2013-05-12	Royal Challengers Bangalore	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	

city	id	season	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	w
Sharjah	459	2014	2014-04-17	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Royal Challengers Bangalore	
Visakhapatnam	313	2012	2012-04-07	Chennai Super Kings	Deccan Chargers	Deccan Chargers	field	normal	0	Chennai Super Kings	

In [204]:

```
# when match held in particular city return index
cities.groups
```

Out[204]:

{'Abu Dhabi': [457, 459, 460, 464, 470, 471, 475], 'Ahmedabad': [179, 186, 194, 197, 481, 486, 497, 500, 525, 532, 535, 539], 'Bangalore': [4, 11, 16, 29, 41, 44, 56, 57, 59, 69, 73, 83, 89, 103, 110, 180, 183, 191, 196, 213, 216, 225, 241, 267, 279, 282, 290, 301, 311, 316, 325, 328, 349, 355, 366, 378, 382, 391, 400, 406, 410, 428, 431, 451, 480, 487, 491, 494, 509, 516, 524, 533, 537, 545, 549, 556, 570, 579, 586, 605, 610, 616, 619, 625, 632, 635], 'Bloemfontein': [160, 164], 'Cape Town': [117, 118, 119, 120, 123, 125, 129], 'Centurion': [132, 135, 136, 146, 147, 148, 155, 156, 165, 169, 170, 171], 'Chandigarh': [31, 34, 45, 47, 60, 68, 72, 80, 93, 96, 107, 176, 192, 195, 204, 214, 242, 255, 283, 286, 320, 329, 332, 339, 353, 365, 390, 399, 408, 435, 439, 447, 504, 508, 511, 543, 551, 565, 568, 578, 585, 588, 596, 611, 614, 621], 'Chennai': [66, 70, 78, 84, 90, 106, 109, 178, 189, 201, 205, 210, 221, 223, 234, 247, 261, 271, 275, 288, 296, 305, 306, 307, 319, 331, 333, 342, 346, 351, 363, 379, 380, 385, 395, 398, 409, 413, 417, 424, 445, 518, 520, 541, 546, 553, 558, 562], 'Cuttack': [185, 188, 336, 347, 485, 490, 496], 'Delhi': [14, 17, 38, 40, 43, 50, 54, 61, 75, 86, 97, 99, 108, 182, 184, 199, 202, 208, 217, 228, 237, 252, 258, 262, 265, 273, 300, 317, 322, 334, 341, 344, 356, 368, 371, 384, 393, 403, 407, 437, 442, 453, 454, 479, 482, 484, 488, 501, 522, 534, 538, 542, 547, 582, 592, 598, 601, 608, 633, 634], 'Dharamsala': [224, 227, 292, 295, 299, 370, 373, 411, 449], 'Durban': [122, 124, 126, 127, 130, 133, 134, 138, 144, 145, 157, 158, 159, 167, 168], 'East London': [137, 143, 149], 'Hyderabad': [0, 5, 18, 20, 35, 42, 46, 58, 65, 67, 76, 79, 91, 101, 113, 235, 244, 248, 259, 274, 278, 285, 330, 358, 372, 375, 383, 387, 404, 422, 427, 434, 448, 452, 492, 495, 499, 502, 550, 563, 567, 571, 583, 587, 593, 597, 602, 609, 617], 'Indore': [3, 7, 21, 289, 293], 'Jaipur': [64, 77, 82, 88, 92, 100, 112, 211, 218, 222, 240, 245, 260, 266, 270, 284, 287, 310, 313, 327, 337, 348, 360, 364, 376, 388, 397, 402, 415, 419, 430, 432, 441], 'Johannesburg': [140, 142, 153, 162, 163, 166, 172, 173], 'Kanpur': [48, 51, 626, 629], 'Kimberley': [150, 151, 154], 'Kochi': [236, 251, 264, 268, 277], 'Kolkata': [10, 13, 22, 26, 30, 39, 52, 63, 74, 87, 94, 102, 104, 111, 177, 181, 203, 207, 212, 226, 229, 239, 250, 254, 257, 269, 280, 302, 308, 321, 324, 343, 352, 362, 367, 381, 396, 405, 412, 414, 426, 455, 456, 503, 505, 510, 513, 517, 521, 526, 544, 554, 559, 575, 577, 580, 607, 613, 620, 623, 630], 'Mumbai': [6, 9, 15, 23, 27, 36, 49, 55, 62, 71, 81, 85, 95, 98, 105, 114, 115, 116, 174, 175, 187, 190, 193, 198, 200, 206, 220, 230, 231, 232, 233, 238, 243, 246, 249, 253, 256, 263, 272, 276, 281, 291, 294, 297, 298, 303, 304, 309, 318, 326, 335, 345, 354, 359, 369, 389, 394, 416, 420, 429, 433, 443, 446, 478, 483, 489, 507, 512, 514, 515, 523, 529, 540, 548, 552, 555, 557, 561, 566, 569, 572, 576, 584, 589, 599], 'Nagpur': [209, 215, 219], 'Port Elizabeth': [121, 128, 131, 139, 141, 152, 161], 'Pune': [1, 8, 24, 28, 32, 37, 53, 314, 323, 338, 340, 350, 357, 361, 374, 377, 386, 392, 401, 421, 425, 436, 438, 450, 519, 527, 531, 573, 591, 595, 600, 604], 'Raipur': [418, 423, 560, 564, 627, 631], 'Rajkot': [2, 12, 19, 25, 33, 581, 590, 594, 603, 606], 'Ranchi': [440, 444, 477, 493, 498, 506, 574], 'Sharjah': [458, 463, 465, 467, 472, 473], 'Visakhapatnam': [312, 315, 528, 530, 536, 612, 615, 618, 622, 624, 628]}

In [205]:

```
df.iloc[475,:]
```

Out[205]:

```
id                             476
season                        2014
city                      Abu Dhabi
date                    2014-04-29
team1              Rajasthan Royals
team2          Kolkata Knight Riders
toss_winner        Rajasthan Royals
toss_decision                  bat
result                         tie
```

```
dl_applied                        0
winner            Rajasthan Royals
win_by_runs                       0
win_by_wickets                    0
player_of_match         JP Faulkner
venue          Sheikh Zayed Stadium
umpire1                   Aleem Dar
umpire2                AK Chaudhary
umpire3                         NaN
Name: 475, dtype: object
```

get_group()

```
# In all cities we only know about Hyderabad
cities.get_group("Hyderabad")
```

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	v
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	
5	6	2017	Hyderabad	2017-04-09	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	
18	19	2017	Hyderabad	2017-04-17	Sunrisers Hyderabad	Kings XI Punjab	Kings XI Punjab	field	normal	0	Sunrisers Hyderabad	
20	21	2017	Hyderabad	2017-04-19	Sunrisers Hyderabad	Delhi Daredevils	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad	
35	36	2017	Hyderabad	2017-04-30	Sunrisers Hyderabad	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Sunrisers Hyderabad	
42	43	2017	Hyderabad	2017-05-06	Rising Pune Supergiant	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Rising Pune Supergiant	
46	47	2017	Hyderabad	2017-05-08	Mumbai Indians	Sunrisers Hyderabad	Mumbai Indians	bat	normal	0	Sunrisers Hyderabad	
58	59	2017	Hyderabad	2017-05-21	Mumbai Indians	Rising Pune Supergiant	Mumbai Indians	bat	normal	0	Mumbai Indians	
65	66	2008	Hyderabad	2008-04-22	Deccan Chargers	Delhi Daredevils	Deccan Chargers	bat	normal	0	Delhi Daredevils	
67	68	2008	Hyderabad	2008-04-24	Deccan Chargers	Rajasthan Royals	Rajasthan Royals	field	normal	0	Rajasthan Royals	
76	77	2008	Hyderabad	2008-05-01	Deccan Chargers	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	w
79	80	2008	Hyderabad	2008-05-25	Deccan Chargers	Royal Challengers Bangalore	Deccan Chargers	bat	normal	0	Royal Challengers Bangalore	
91	92	2008	Hyderabad	2008-05-11	Kolkata Knight Riders	Deccan Chargers	Kolkata Knight Riders	bat	normal	0	Kolkata Knight Riders	
101	102	2008	Hyderabad	2008-05-18	Mumbai Indians	Deccan Chargers	Deccan Chargers	field	normal	0	Mumbai Indians	
113	114	2008	Hyderabad	2008-05-27	Deccan Chargers	Chennai Super Kings	Deccan Chargers	bat	normal	0	Chennai Super Kings	
235	236	2011	Hyderabad	2011-04-09	Deccan Chargers	Rajasthan Royals	Rajasthan Royals	field	normal	0	Rajasthan Royals	
244	245	2011	Hyderabad	2011-04-14	Deccan Chargers	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Deccan Chargers	
248	249	2011	Hyderabad	2011-04-16	Deccan Chargers	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	
259	260	2011	Hyderabad	2011-04-24	Mumbai Indians	Deccan Chargers	Deccan Chargers	field	normal	0	Mumbai Indians	
274	275	2011	Hyderabad	2011-05-03	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	field	normal	0	Kolkata Knight Riders	
278	279	2011	Hyderabad	2011-05-05	Deccan Chargers	Delhi Daredevils	Delhi Daredevils	field	normal	0	Delhi Daredevils	
285	286	2011	Hyderabad	2011-05-10	Deccan Chargers	Pune Warriors	Deccan Chargers	bat	normal	0	Pune Warriors	
330	331	2012	Hyderabad	2012-05-10	Deccan Chargers	Delhi Daredevils	Deccan Chargers	bat	normal	0	Delhi Daredevils	
358	359	2012	Hyderabad	2012-05-08	Kings XI Punjab	Deccan Chargers	Deccan Chargers	field	normal	0	Kings XI Punjab	
372	373	2012	Hyderabad	2012-05-18	Rajasthan Royals	Deccan Chargers	Rajasthan Royals	bat	normal	0	Deccan Chargers	
375	376	2012	Hyderabad	2012-05-20	Deccan Chargers	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Deccan Chargers	
383	384	2013	Hyderabad	2013-04-05	Sunrisers Hyderabad	Pune Warriors	Pune Warriors	field	normal	0	Sunrisers Hyderabad	

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	w
387	388	2013	Hyderabad	2013-04-07	Royal Challengers Bangalore	Sunrisers Hyderabad	Royal Challengers Bangalore	bat	tie	0	Sunrisers Hyderabad	
404	405	2013	Hyderabad	2013-04-19	Kings XI Punjab	Sunrisers Hyderabad	Kings XI Punjab	bat	normal	0	Sunrisers Hyderabad	
422	423	2013	Hyderabad	2013-05-01	Mumbai Indians	Sunrisers Hyderabad	Mumbai Indians	bat	normal	0	Sunrisers Hyderabad	
427	428	2013	Hyderabad	2013-05-04	Delhi Daredevils	Sunrisers Hyderabad	Delhi Daredevils	bat	normal	0	Sunrisers Hyderabad	
434	435	2013	Hyderabad	2013-05-08	Chennai Super Kings	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Chennai Super Kings	
448	449	2013	Hyderabad	2013-05-17	Sunrisers Hyderabad	Rajasthan Royals	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad	
452	453	2013	Hyderabad	2013-05-19	Kolkata Knight Riders	Sunrisers Hyderabad	Kolkata Knight Riders	bat	normal	0	Sunrisers Hyderabad	
492	493	2014	Hyderabad	2014-05-12	Sunrisers Hyderabad	Mumbai Indians	Sunrisers Hyderabad	bat	normal	0	Mumbai Indians	
495	496	2014	Hyderabad	2014-05-14	Sunrisers Hyderabad	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	
499	500	2014	Hyderabad	2014-05-18	Sunrisers Hyderabad	Kolkata Knight Riders	Sunrisers Hyderabad	bat	normal	0	Kolkata Knight Riders	
502	503	2014	Hyderabad	2014-05-20	Royal Challengers Bangalore	Sunrisers Hyderabad	Royal Challengers Bangalore	bat	normal	0	Sunrisers Hyderabad	
550	551	2015	Hyderabad	2015-05-02	Sunrisers Hyderabad	Chennai Super Kings	Chennai Super Kings	field	normal	0	Sunrisers Hyderabad	
563	564	2015	Hyderabad	2015-05-11	Sunrisers Hyderabad	Kings XI Punjab	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad	
567	568	2015	Hyderabad	2015-05-15	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	bat	normal	1	Royal Challengers Bangalore	
571	572	2015	Hyderabad	2015-05-17	Sunrisers Hyderabad	Mumbai Indians	Sunrisers Hyderabad	bat	normal	0	Mumbai Indians	
583	584	2016	Hyderabad	2016-04-16	Sunrisers Hyderabad	Kolkata Knight Riders	Sunrisers Hyderabad	bat	normal	0	Kolkata Knight Riders	
587	588	2016	Hyderabad	2016-04-18	Mumbai Indians	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	

id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	v
593	594	2016	Hyderabad	2016-04-23	Kings XI Punjab	Sunrisers Hyderabad	Sunrisers Hyderabad		field	normal	0	Sunrisers Hyderabad	
597	598	2016	Hyderabad	2016-04-26	Sunrisers Hyderabad	Rising Pune Supergiants	Rising Pune Supergiants		field	normal	1	Rising Pune Supergiants	
602	603	2016	Hyderabad	2016-04-30	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore		field	normal	0	Sunrisers Hyderabad	
609	610	2016	Hyderabad	2016-05-06	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad		field	normal	0	Sunrisers Hyderabad	
617	618	2016	Hyderabad	2016-05-12	Sunrisers Hyderabad	Delhi Daredevils	Delhi Daredevils		field	normal	0	Delhi Daredevils	

In [209]:

```
cities.get_group("Hyderabad").shape[0]
```

Out[209]:

49

In [218]:

```
cities.sum()
```

Out[218]:

city	id	season	dl_applied	win_by_runs	win_by_wickets	umpire3
Abu Dhabi	3263	14098	0	157	16	0.0
Ahmedabad	4863	24156	0	180	38	0.0
Bangalore	22299	132861	3	1263	232	0.0
Bloemfontein	326	4018	0	14	6	0.0
Cape Town	858	14063	1	145	18	0.0
Centurion	1842	24108	1	81	49	0.0
Chandigarh	15325	92582	0	576	167	0.0
Chennai	15096	96554	0	715	111	0.0
Cuttack	2534	14086	0	73	20	0.0
Delhi	19600	120771	2	876	192	0.0
Dharamsala	2949	18103	0	279	23	0.0
Durban	2147	30135	1	159	33	0.0
East London	432	6027	0	87	7	0.0
Hyderabad	15907	98649	2	486	184	0.0
Indore	618	10073	0	0	36	0.0
Jaipur	9210	66363	0	460	151	0.0
Johannesburg	1279	16072	0	25	35	0.0
Kanpur	1358	8066	0	0	22	0.0

city	id	season	dl_applied	win_by_runs	win_by_wickets	umpire3
Kimberley	458	6027	0	53	10	0.0
Kochi	1301	10055	1	110	13	0.0
Kolkata	19973	122783	3	587	224	0.0
Mumbai	26283	171049	0	1102	264	0.0
Nagpur	646	6030	0	15	6	0.0
Port Elizabeth	980	14063	0	200	22	0.0
Pune	11238	64455	0	545	98	0.0
Raipur	3229	12088	0	21	25	0.0
Rajkot	3075	20165	0	70	48	0.0
Ranchi	3439	14097	0	41	24	0.0
Sharjah	2804	12084	0	74	26	0.0
Visakhapatnam	5951	22165	2	282	22	0.0

In [223]:

```python
cities.sum()["win_by_runs"].sort_values(ascending=False).plot(kind="pie")
```

Out[223]:

```
<AxesSubplot:ylabel='win_by_runs'>
```
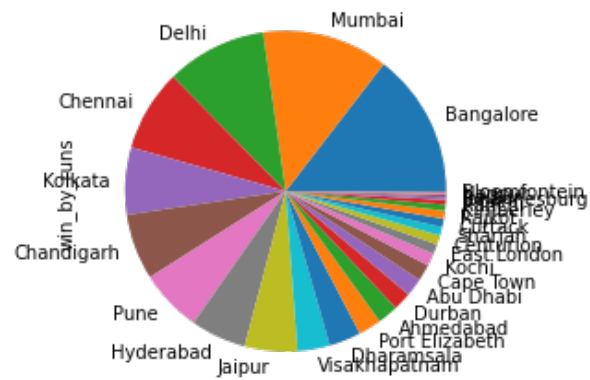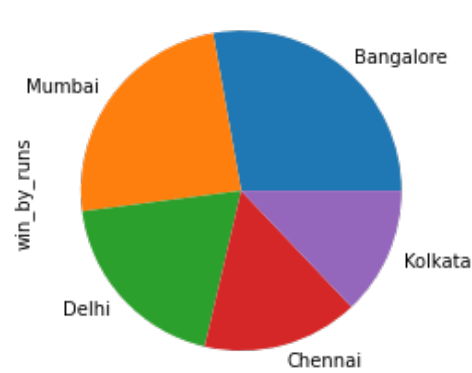

In [225]:

```python
(cities.sum()["win_by_runs"].sort_values(ascending=False).head()).plot(kind="pie")
```

Out[225]:

```
<AxesSubplot:ylabel='win_by_runs'>
```


In [254]:

```python
df.groupby("city")["winner"].count().sort_values(ascending=False)
```

Out[254]:

```
city
Mumbai          85
```

```
Bangalore              64
Kolkata                61
Delhi                  59
Hyderabad              49
Chennai                48
Chandigarh             46
Jaipur                 33
Pune                   32
Durban                 15
Centurion              12
Ahmedabad              12
Visakhapatnam          11
Rajkot                 10
Dharamsala              9
Johannesburg            8
Cape Town               7
Cuttack                 7
Ranchi                  7
Port Elizabeth          7
Abu Dhabi               7
Raipur                  6
Sharjah                 6
Indore                  5
Kochi                   5
Kanpur                  4
Nagpur                  3
Kimberley               3
East London             3
Bloemfontein            2
Name: winner, dtype: int64
```

In [261]:

```
df
```

Out[261]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	v
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	
...	...	...	...	...	...	...	...	...	...	...	...	...
631	632	2016	Raipur	2016-05-22	Delhi Daredevils	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Royal Challengers Bangalore	
632	633	2016	Bangalore	2016-05-24	Gujarat Lions	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Royal Challengers Bangalore	
633	634	2016	Delhi	2016-05-25	Sunrisers Hyderabad	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Sunrisers Hyderabad	

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	v
634	635	2016	Delhi	2016-05-27	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	
635	636	2016	Bangalore	2016-05-29	Sunrisers Hyderabad	Royal Challengers Bangalore	Sunrisers Hyderabad	bat	normal	0	Sunrisers Hyderabad	

636 rows × 18 columns
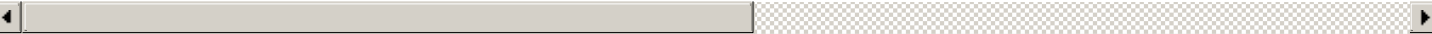
In [263]:

```
# In every years winner city
# 2010 : Hyderabad
```

In [272]:

```
#df[["city","winner"]]
df.drop_duplicates(subset="season",keep="last")[["season","winner"]].sort_values("season"
)
```

Out[272]:

	season	winner
116	2008	Rajasthan Royals
173	2009	Deccan Chargers
233	2010	Chennai Super Kings
306	2011	Chennai Super Kings
380	2012	Kolkata Knight Riders
456	2013	Mumbai Indians
516	2014	Kolkata Knight Riders
575	2015	Mumbai Indians
635	2016	Sunrisers Hyderabad
58	2017	Mumbai Indians

Merge two dataset

In [283]:

```
! ls
```

```
 __MACOSX    so_survey_2019.pdf       Untitled1.html
'matches - matches.csv'    survey_results_public.csv    Untitled1.ipynb
 README_2019.txt    survey_results_schema.csv    Untitled.ipynb
```

In [284]:

```
survey = pd.read_csv("survey_results_public.csv")
```

In [285]:

```
survey.shape
```

Out[285]:

```
(88883, 85)
```

In [286]:

```
match =  pd.read_csv("matches - matches.csv")
```

In [287]:

```
match.shape
```

Out[287]:

```
(636, 18)
```

In [289]:

```
survey.head(1)
```

Out[289]:

	Respondent	MainBranch	Hobbyist	OpenSourcer	OpenSource	Employment	Country	Student	EdLevel	Unde
0	1	I am a student who is learning to code	Yes	Never	The quality of OSS and closed source software ...	Not employed, and not looking for work	United Kingdom	No	Primary/elementary school	

1 rows × 85 columns

◄ |████████████████████████| ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒ | ►

In [291]:

```
match.head(1)
```

Out[291]:

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

◄ |████████████████████████| ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒ | ►

In [302]:

```
new = survey.merge(match,left_on="Respondent",right_on="id")
```

In [306]:

```
new.shape
```

Out[306]:

```
(636, 103)
```

In [305]:

```
new.head(1)
```

Out[305]:

	Respondent	MainBranch	Hobbyist	OpenSourcer	OpenSource	Employment	Country	Student	EdLevel	Unde
0	1	I am a student who is learning to code	Yes	Never	The quality of OSS and closed source software ...	Not employed, and not looking for work	United Kingdom	No	Primary/elementary school	

◄ |████████| ▒▒▒ | ►

In [301]:

```
pd.set_option("display.max_columns",103)
```

In [308]:

```
# new
```

Pivot Table

In [320]:

```
df.head(1)
```

Out[320]:

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

In [333]:

```
df.pivot_table(index="season",columns="winner",values="win_by_runs",aggfunc="count")
```

Out[333]:

winner	Chennai Super Kings	Deccan Chargers	Delhi Daredevils	Gujarat Lions	Kings XI Punjab	Kochi Tuskers Kerala	Kolkata Knight Riders	Mumbai Indians	Pune Warriors	Rajasthan Royals	Rising Pune Supergiant	Rising Pune Supergi
season												
2008	9.0	2.0	7.0	NaN	10.0	NaN	6.0	7.0	NaN	13.0	NaN	
2009	8.0	9.0	10.0	NaN	7.0	NaN	3.0	5.0	NaN	6.0	NaN	
2010	9.0	8.0	7.0	NaN	4.0	NaN	7.0	11.0	NaN	6.0	NaN	
2011	11.0	6.0	4.0	NaN	7.0	6.0	8.0	10.0	4.0	6.0	NaN	
2012	10.0	4.0	11.0	NaN	8.0	NaN	12.0	10.0	4.0	7.0	NaN	
2013	12.0	NaN	3.0	NaN	8.0	NaN	6.0	13.0	4.0	11.0	NaN	
2014	10.0	NaN	2.0	NaN	12.0	NaN	11.0	7.0	NaN	7.0	NaN	
2015	10.0	NaN	5.0	NaN	3.0	NaN	7.0	10.0	NaN	7.0	NaN	
2016	NaN	NaN	7.0	9.0	4.0	NaN	8.0	7.0	NaN	NaN	NaN	
2017	NaN	NaN	6.0	4.0	7.0	NaN	9.0	12.0	NaN	NaN	10.0	

corelation

In [336]:

```
df.corr()
```

Out[336]:

	id	season	dl_applied	win_by_runs	win_by_wickets	umpire3
id	1.000000	0.471087	0.024281	-0.010263	-0.015510	NaN
season	0.471087	1.000000	0.004170	-0.016815	-0.000708	NaN
dl_applied	0.024281	0.004170	1.000000	-0.010893	-0.011640	NaN

		id	season	dl_applied	win_by_runs	win_by_wickets	umpire3
win_by_runs	-0.010263	-0.016815	-0.010893	1.000000	-0.565181	NaN
win_by_wickets	-0.015510	-0.000708	-0.011640	-0.565181	1.000000	NaN
umpire3	NaN	NaN	NaN	NaN	NaN	NaN

In [342]:

```
df.corrwith(match)
```

Out[342]:

```
id               1.0
season           1.0
dl_applied       1.0
win_by_runs      1.0
win_by_wickets   1.0
umpire3          NaN
dtype: float64
```

In [343]:

```
import seaborn as sns
```

In [345]:

```
sns.heatmap(df.corr())
```

Out[345]:

```
<AxesSubplot:>
```
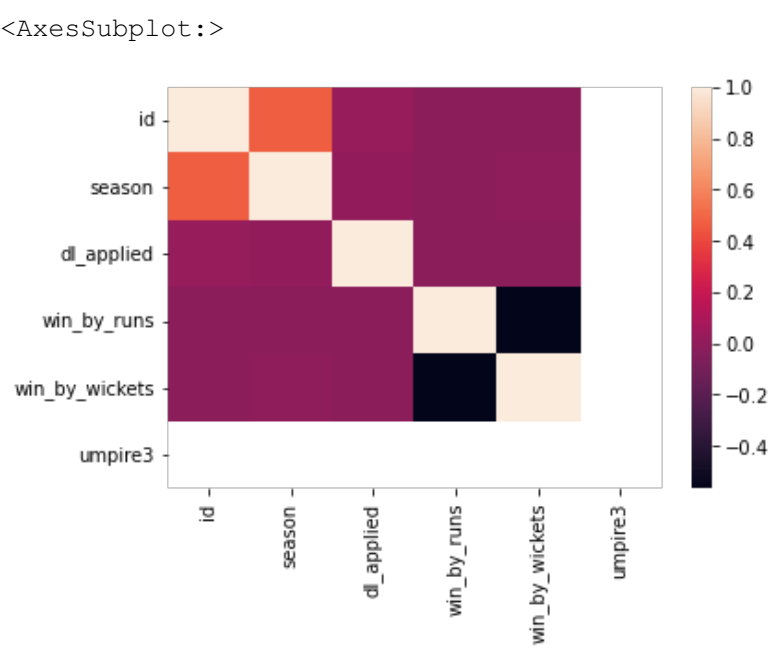

Rename Column

In [346]:

```
mt=match
```

In [349]:

```
mt.head(1)
```

Out[349]:

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by
0	1	2017	Hyderabad	2017-	Sunrisers	Royal Challengers	Royal Challengers	field	normal	0	Sunrisers	

id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by
1	2017	Hyderabad	04-05	Hyderabad	Challengers	Bangalore	field	normal		Hyderabad	

```
# Change the date name to dt and city name to place.
```

```
mt.rename(columns={"city":"place","date":"dt"},inplace=True)
```

```
mt.head(1)
```

Out[363]:

	id	season	place	dt	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	

```
match.head(1)
```

Out[366]:

	id	season	place	dt	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	

Set_index and reset_index

---

reset_index is also used to convert series data Dataframe

```
df.head(1)
```

Out[374]:

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

```
# Here pandas give own index by default.
df.set_index("id",inplace=True)
```

```
df.head(1)
```

Out[407]:

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

In [408]:

```
df.reset_index(inplace=True)
```

In [409]:

```
df.head(1)
```

Out[409]:

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

reset_index also used to convert series dataset to DataFrame

In [464]:

```
df["city"]
```

Out[464]:

```
0        Hyderabad
1             Pune
2           Rajkot
3           Indore
4        Bangalore
           ...    
631         Raipur
632      Bangalore
633          Delhi
634          Delhi
635      Bangalore
Name: city, Length: 636, dtype: object
```

In [468]:

```
df["city"].reset_index()
#df["city"].reset_index().drop(columns="index")
```

Out[468]:

	index	city
0	0	Hyderabad
1	1	Pune
2	2	Rajkot
3	3	Indore

	index	city
4	4	Bangalore
...	...	...
631	631	Raipur
632	632	Bangalore
633	633	Delhi
634	634	Delhi
635	635	Bangalore

636 rows × 2 columns

In [410]:

```python
# df.drop(columns="index",inplace=True)
```

Handle Missing Values

1. Drop missing values using dropna()

In [411]:

```python
# In 6 columns there is null values
df.isnull().sum()
```

Out[411]:

```
id                 0
season             0
city               7
date               0
team1              0
team2              0
toss_winner        0
toss_decision      0
result             0
dl_applied         0
winner             3
win_by_runs        0
win_by_wickets     0
player_of_match    3
venue              0
umpire1            1
umpire2            1
umpire3          636
dtype: int64
```

In [412]:

```python
df.shape
```

Out[412]:

```
(636, 18)
```

In [427]:

```python
# If any column you find any NAN values just delete all column
df.dropna(axis=1).shape
```

Out[427]:

```
(636, 12)
```

In [440]:

```python
# Find any row with NAN values remove that whole row
```

```
df.dropna(subset=["umpire1"]).shape
```

Out[440]:

(635, 18)

2.Filling Missing Values With fillna()

In [442]:

```
df.drop(columns="umpire3",inplace= True)
```

In [446]:

```
df.head(1)
```

Out[446]:

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

In [452]:

```
df.isnull().sum().sort_values(ascending=False)
```

Out[452]:

```
city              7
player_of_match   3
winner            3
umpire2           1
umpire1           1
dl_applied        0
venue             0
win_by_wickets    0
win_by_runs       0
id                0
season            0
toss_decision     0
toss_winner       0
team2             0
team1             0
date              0
result            0
dtype: int64
```

In [453]:

```
df["season"].mean()
```

Out[453]:

2012.4905660377358

In [454]:

```
# All NAN values fill with some values
df["season"].fillna(df["season"].mean())
```

Out[454]:

```
0     2017
1     2017
2     2017
3     2017
4     2017
```

```
        ...
631    2016
632    2016
633    2016
634    2016
635    2016
Name: season, Length: 636, dtype: int64
```

In [ ]: