

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: ipl_df = pd.read_csv(r"IPL Matches.csv")
```

```
In [3]: ipl_df
```

```
Out[3]:
```

	id	city	date	player_of_match	venue	neutral_venue	team1
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab
2	335984	Delhi	2008-04-19	MF Maharooof	Feroz Shah Kotla	0	Delhi Daredevils
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders
...
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0	Royal Challengers Bangalore
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mumbai Indians
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0	Royal Challengers Bangalore
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0	Delhi Capitals
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0	Delhi Capitals

816 rows × 17 columns

```
In [4]: ipl_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
```

```
In [5]: ipl_df.describe(include="all")
```

```
Out[5]:
```

	id	city	date	player_of_match	venue	neutral_venue	team1
count	8.160000e+02	803	816	812	816	816.000000	816
unique	NaN	32	596	233	36	NaN	15
top	NaN	Mumbai	2014-05-11	AB de Villiers	Eden Gardens	NaN	Royal Challengers Bangalore
freq	NaN	101	2	23	77	NaN	108
mean	7.563496e+05	NaN	NaN	NaN	NaN	0.094363	NaN
std	3.058943e+05	NaN	NaN	NaN	NaN	0.292512	NaN
min	3.359820e+05	NaN	NaN	NaN	NaN	0.000000	NaN
25%	5.012278e+05	NaN	NaN	NaN	NaN	0.000000	NaN
50%	7.292980e+05	NaN	NaN	NaN	NaN	0.000000	NaN
75%	1.082626e+06	NaN	NaN	NaN	NaN	0.000000	NaN
max	1.237181e+06	NaN	NaN	NaN	NaN	1.000000	NaN

Finding Missing Values

```
In [7]: ipl_df.isna().sum()
```

```
Out[7]: 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
```

```
In [8]: (ipl_df.isna().sum()/len(ipl_df))*100
```

```
Out[8]: id                0.000000
        city              1.593137
        date              0.000000
        player_of_match   0.490196
        venue             0.000000
        neutral_venue     0.000000
        team1             0.000000
        team2             0.000000
        toss_winner       0.000000
        toss_decision     0.000000
        winner            0.490196
        result            0.490196
        result_margin     2.083333
        eliminator        0.490196
        method           97.671569
        umpire1           0.000000
        umpire2           0.000000
        dtype: float64
```

Checking the shape and size of dataset

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

```
Out[10]: (816, 17)
```

```
In [11]: ipl_df.size
```

Out[11]: 13872

This Dataset has 816 rows and 17 columns and in total 13872 datapoints including null and non null
noth

```
In [13]: for i,j in enumerate(ipl_df.columns):  
         if i == 0:  
             print("Name of Columns: ")  
             print(f"{i+1}. {j}")
```

Name of Columns:
1. id
2. city
3. date
4. player_of_match
5. venue
6. neutral_venue
7. team1
8. team2
9. toss_winner
10. toss_decision
11. winner
12. result
13. result_margin
14. eliminator
15. method
16. umpire1
17. umpire2

The dropna mindset

```
In [15]: ipl_non_na = ipl_df.copy()
```

```
In [16]: ipl_non_na["method"].value_counts()
```

```
Out[16]: method  
D/L      19  
Name: count, dtype: int64
```

```
In [17]: ipl_non_na["method"].unique()
```

```
Out[17]: array([nan, 'D/L'], dtype=object)
```

```
In [18]: ipl_nonna_df = ipl_non_na.fillna({"method": "Fair Match"})
```

```
In [19]: ipl_nonna_df
```

Out[19]:

	id	city	date	player_of_match	venue	neutral_venue	team1
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab
2	335984	Delhi	2008-04-19	MF Maharoo	Feroz Shah Kotla	0	Delhi Daredevils
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders
...
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0	Royal Challengers Bangalore
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mumbai Indians
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0	Royal Challengers Bangalore
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0	Delhi Capitals
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0	Delhi Capitals

816 rows × 17 columns

fillna - The `fillna()` method replaces the NULL values with a specified value. The `fillna()` method returns a new DataFrame object unless the `inplace` parameter is set to `True`, in that case the `fillna()` method does the replacing in the original DataFrame instead.

```
In [21]: ipl_nonna_df.isna().sum()
```

```
Out[21]: 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              0
umpire1             0
umpire2             0
dtype: int64
```

```
In [22]: ipl_nonna_df.dropna(inplace=True)
```

```
In [23]: ipl_nonna_df.isna().sum()
```

```
Out[23]: id                0
city                0
date                0
player_of_match     0
venue               0
neutral_venue       0
team1               0
team2               0
toss_winner         0
toss_decision       0
winner              0
result              0
result_margin       0
eliminator          0
method              0
umpire1             0
umpire2             0
dtype: int64
```

Locations for matches

```
In [25]: for i,j in enumerate(ipl_nonna_df["city"].unique()):  
         if i == 0:  
             print("Location List")  
         print(f"{i + 1}. {j}")
```

```
Location List  
1. Bangalore  
2. Chandigarh  
3. Delhi  
4. Mumbai  
5. Kolkata  
6. Jaipur  
7. Hyderabad  
8. Chennai  
9. Cape Town  
10. Port Elizabeth  
11. Durban  
12. Centurion  
13. East London  
14. Johannesburg  
15. Kimberley  
16. Bloemfontein  
17. Ahmedabad  
18. Cuttack  
19. Nagpur  
20. Dharamsala  
21. Kochi  
22. Indore  
23. Visakhapatnam  
24. Pune  
25. Raipur  
26. Ranchi  
27. Abu Dhabi  
28. Rajkot  
29. Kanpur  
30. Bengaluru  
31. Sharjah  
32. Dubai
```

Thus in total we have a list of 32 locations for ipl matches with Bangalore and Bengaluru being repeated so in total 31 only

```
In [27]: ipl_nonna_df.columns
```

```
Out[27]: 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')
```

```
In [28]: ipl_nonna_df.describe(include="all")
```

Out[28]:

	id	city	date	player_of_match	venue	neutral_venue	team1
count	7.860000e+02	786	786	786	786	786.000000	786
unique	NaN	32	578	232	36	NaN	15
top	NaN	Mumbai	2014-05-13	AB de Villiers	Eden Gardens	NaN	Royal Challengers Bangalore
freq	NaN	100	2	22	77	NaN	10
mean	7.538069e+05	NaN	NaN	NaN	NaN	0.078880	NaN
std	3.078221e+05	NaN	NaN	NaN	NaN	0.269724	NaN
min	3.359820e+05	NaN	NaN	NaN	NaN	0.000000	NaN
25%	5.012222e+05	NaN	NaN	NaN	NaN	0.000000	NaN
50%	6.636760e+05	NaN	NaN	NaN	NaN	0.000000	NaN
75%	1.082627e+06	NaN	NaN	NaN	NaN	0.000000	NaN
max	1.237181e+06	NaN	NaN	NaN	NaN	1.000000	NaN

```
In [29]: ipl_nonna_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 786 entries, 0 to 815
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    786 non-null   int64
1   city                  786 non-null   object
2   date                  786 non-null   object
3   player_of_match       786 non-null   object
4   venue                 786 non-null   object
5   neutral_venue         786 non-null   int64
6   team1                 786 non-null   object
7   team2                 786 non-null   object
8   toss_winner           786 non-null   object
9   toss_decision         786 non-null   object
10  winner                786 non-null   object
11  result                786 non-null   object
12  result_margin         786 non-null   float64
13  eliminator            786 non-null   object
14  method                786 non-null   object
15  umpire1               786 non-null   object
16  umpire2               786 non-null   object
dtypes: float64(1), int64(2), object(14)
memory usage: 110.5+ KB
```

```
In [30]: ipl_nonna_df["date"] = pd.to_datetime(ipl_nonna_df["date"])
```



```
In [31]: ipl_nonna_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 786 entries, 0 to 815
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                     786 non-null    int64
1   city                   786 non-null    object
2   date                   786 non-null    datetime64[ns]
3   player_of_match        786 non-null    object
4   venue                  786 non-null    object
5   neutral_venue          786 non-null    int64
6   team1                  786 non-null    object
7   team2                  786 non-null    object
8   toss_winner            786 non-null    object
9   toss_decision          786 non-null    object
10  winner                 786 non-null    object
11  result                 786 non-null    object
12  result_margin          786 non-null    float64
13  eliminator             786 non-null    object
14  method                 786 non-null    object
15  umpire1                786 non-null    object
16  umpire2                786 non-null    object
dtypes: datetime64[ns](1), float64(1), int64(2), object(13)
memory usage: 110.5+ KB
```

```
In [32]: ipl_nonna_df.columns
```

```
Out[32]: 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')
```

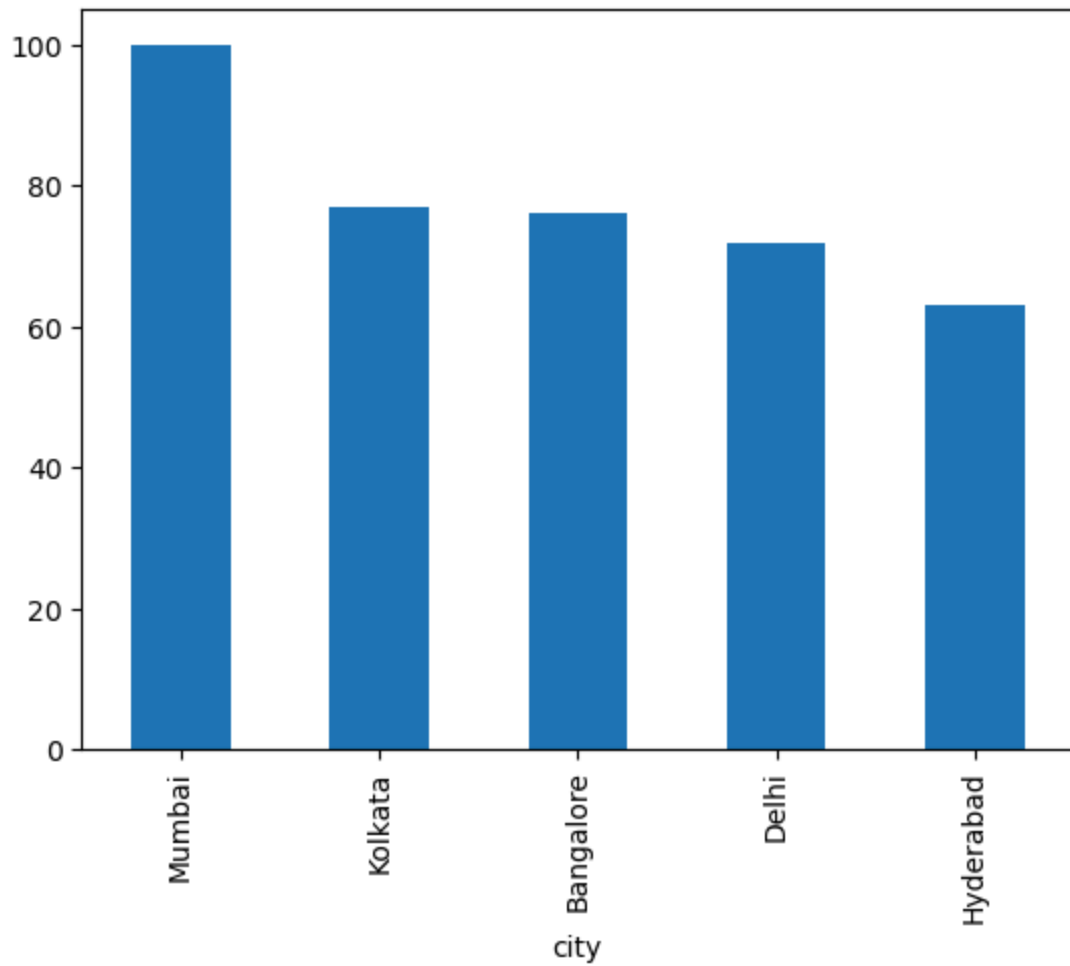
```
In [33]: ipl_nonna_df.replace(to_replace= "Bengaluru", value = "Bangalore",inplace=True )
```

```
In [34]: ipl_nonna_df["city"].value_counts()
```

```
Out[34]: city
Mumbai          100
Kolkata          77
Bangalore        76
Delhi            72
Hyderabad        63
Chandigarh       56
Chennai          56
Jaipur           47
Pune             38
Abu Dhabi        27
Dubai            23
Durban           15
Visakhapatnam    13
Centurion        12
Sharjah          12
Ahmedabad        11
Dharamsala       9
Rajkot           9
Indore           9
Johannesburg     8
Cuttack          7
Port Elizabeth   7
Ranchi           7
Raipur           6
Cape Town        6
Kochi            5
Kanpur           4
Nagpur           3
East London      3
Kimberley        3
Bloemfontein     2
Name: count, dtype: int64
```

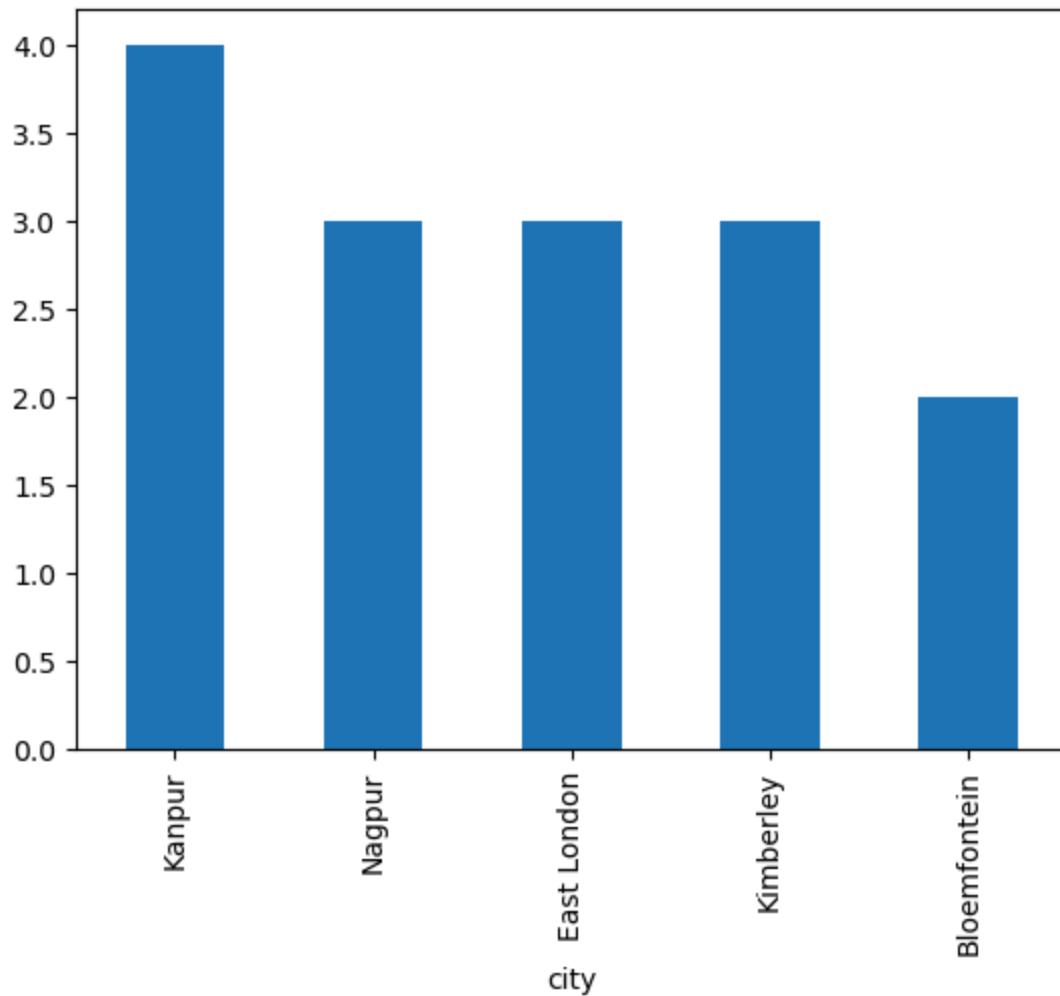
```
In [35]: ipl_nonna_df["city"].value_counts().head(5).plot(kind = "bar")
```

```
Out[35]: <Axes: xlabel='city'>
```



```
In [36]: ipl_nonna_df["city"].value_counts().tail(5).plot(kind = "bar")
```

```
Out[36]: <Axes: xlabel='city'>
```



```
In [37]: ipl_nonna_df["Year"] = ipl_nonna_df["date"].dt.year
```

```
In [38]: ipl_nonna_df.sample()
```

```
Out[38]:
```

	id	city	date	player_of_match	venue	neutral_venue	team1	team2
370	598046	Mumbai	2013-05-05	MG Johnson	Wankhede Stadium	0	Mumbai Indians	Chennai Super Kings

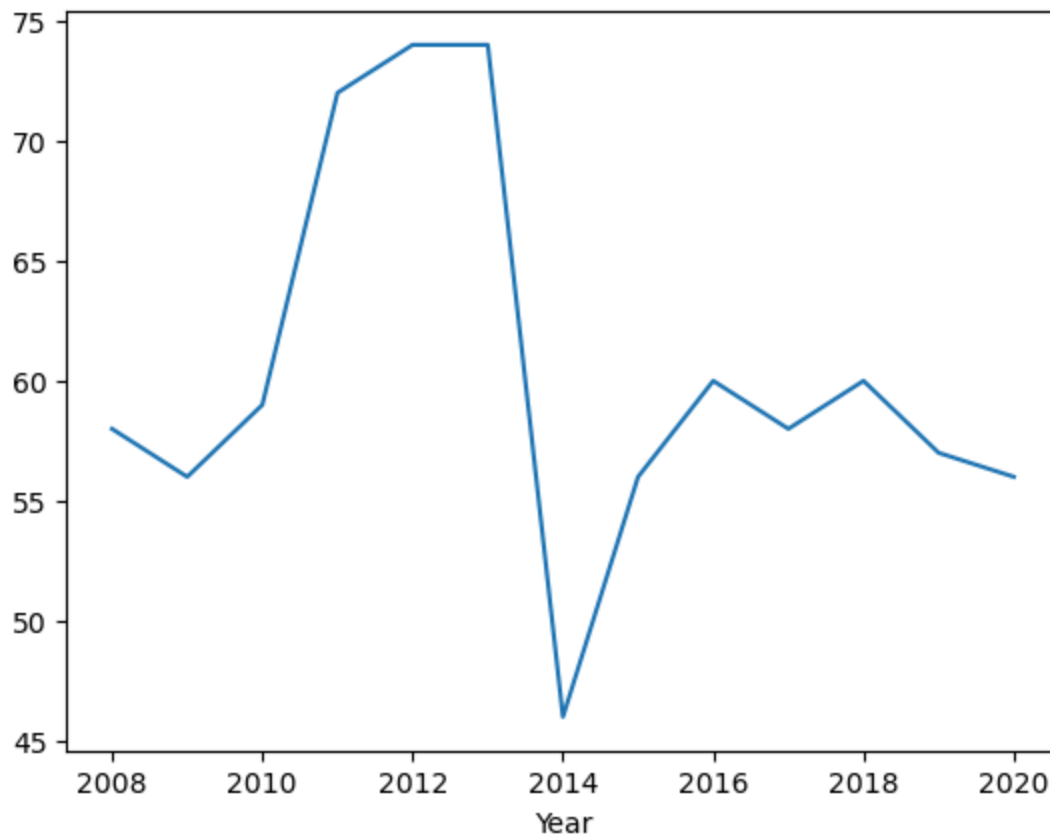
```
In [39]: ipl_nonna_df["Year"].value_counts().sort_values(ascending = False)
```

```
Out[39]: Year
2012    74
2013    74
2011    72
2016    60
2018    60
2010    59
2008    58
2017    58
2019    57
2009    56
2015    56
2020    56
2014    46
Name: count, dtype: int64
```

Find above the total matches by year

```
In [41]: ipl_nonna_df["Year"].value_counts().sort_index().plot()
```

```
Out[41]: <Axes: xlabel='Year'>
```

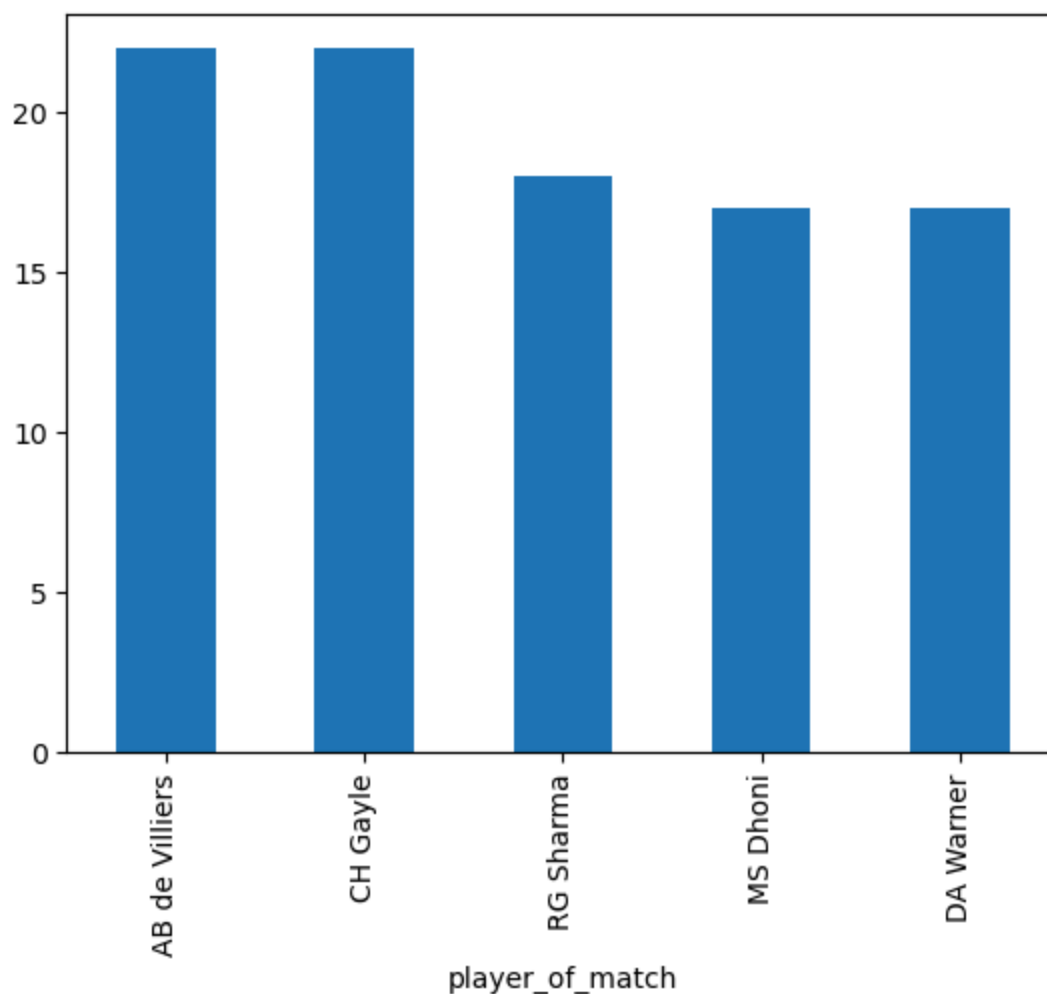


```
In [42]: ipl_nonna_df["player_of_match"].value_counts()
```

```
Out[42]: player_of_match
AB de Villiers    22
CH Gayle          22
RG Sharma         18
MS Dhoni          17
DA Warner         17
..
JEC Franklin      1
Shoaib Akhtar     1
S Aravind          1
BA Bhatt          1
PP Shaw           1
Name: count, Length: 232, dtype: int64
```

```
In [43]: ipl_nonna_df["player_of_match"].value_counts().head(5).plot(kind = "bar")
```

```
Out[43]: <Axes: xlabel='player_of_match'>
```



```
In [44]: players = set(",".join(ipl_nonna_df["player_of_match"].unique()).lower().split(","))
```

```
In [45]: type(players)
```

```
Out[45]: set
```

```
In [46]: len(players)
```

```
Out[46]: 232
```

```
In [47]: ipl_nonna_df["player_of_match"].value_counts()
```

```
Out[47]: player_of_match
AB de Villiers    22
CH Gayle          22
RG Sharma         18
MS Dhoni          17
DA Warner         17
..
JEC Franklin      1
Shoaib Akhtar     1
S Aravind          1
BA Bhatt          1
PP Shaw           1
Name: count, Length: 232, dtype: int64
```

```
In [48]: ipl_nonna_df["player_of_match"].value_counts().head(10)
```

```
Out[48]: player_of_match
AB de Villiers    22
CH Gayle          22
RG Sharma         18
MS Dhoni          17
DA Warner         17
SR Watson         16
YK Pathan         15
SK Raina          14
G Gambhir         13
V Kohli           12
Name: count, dtype: int64
```

Find above the list of top 10 players who have won the most player of the match awards

```
In [50]: ipl_nonna_df["player_of_match"].value_counts().tail(10)
```

```
Out[50]: player_of_match
JD Ryder          1
AD Mascarenhas    1
MN Samuels        1
CRD Fernando      1
RE Levi           1
JEC Franklin      1
Shoaib Akhtar     1
S Aravind          1
BA Bhatt          1
PP Shaw           1
Name: count, dtype: int64
```

```
ipl_nonna_df["venue"].nunique()
```

Out[51]: 36

```
venue = set("|".join(ipl_nonna_df["venue"].unique()).lower().split("|"))
```

```
len(venue)
```

Out[53]: 36

No duplicate venue locations in dataset

```
ipl_nonna_df["venue"].value_counts()
```

```
Out[55]:
```

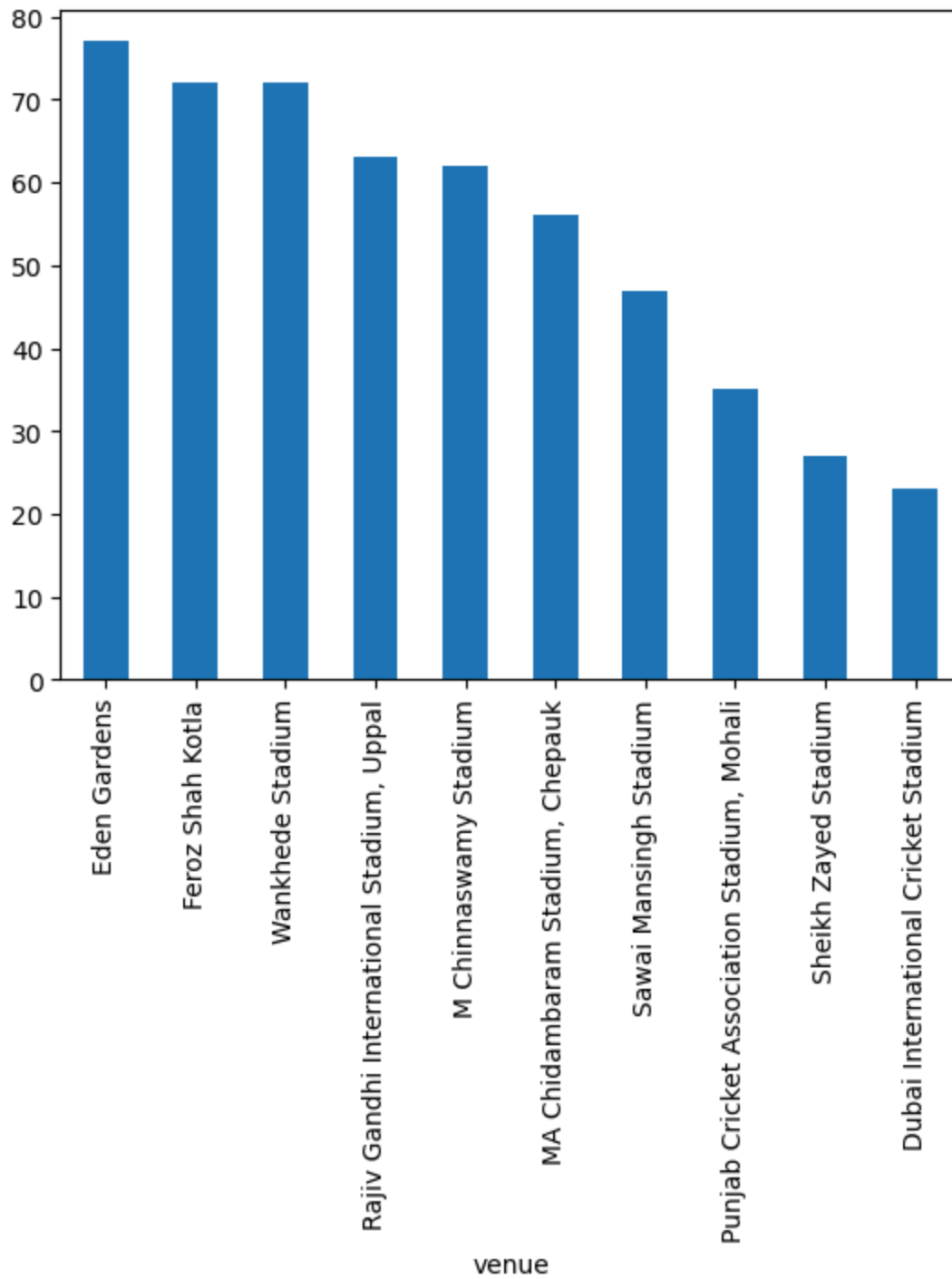
venue	77
Eden Gardens	72
Feroz Shah Kotla	72
Wankhede Stadium	63
Rajiv Gandhi International Stadium, Uppal	62
M Chinnaswamy Stadium	56
MA Chidambaram Stadium, Chepauk	47
Sawai Mansingh Stadium	35
Punjab Cricket Association Stadium, Mohali	27
Sheikh Zayed Stadium	23
Dubai International Cricket Stadium	21
Punjab Cricket Association IS Bindra Stadium, Mohali	21
Maharashtra Cricket Association Stadium	17
Dr DY Patil Sports Academy	17
Subrata Roy Sahara Stadium	15
Kingsmead	14
M.Chinnaswamy Stadium	13
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium	12
SuperSport Park	12
Sharjah Cricket Stadium	11
Sardar Patel Stadium, Motera	11
Brabourne Stadium	9
Saurashtra Cricket Association Stadium	9
Holkar Cricket Stadium	9
Himachal Pradesh Cricket Association Stadium	8
New Wanderers Stadium	7
JSCA International Stadium Complex	7
Barabati Stadium	7
St George's Park	6
Shaheed Veer Narayan Singh International Stadium	6
Newlands	5
Nehru Stadium	4
Green Park	3
Vidarbha Cricket Association Stadium, Jamtha	3
De Beers Diamond Oval	3
Buffalo Park	2
OUTsurance Oval	
Name: count, dtype: int64	


```
In [56]: ipl_nonna_df["venue"].value_counts().head(10)
```

```
Out[56]: venue
Eden Gardens                77
Feroz Shah Kotla            72
Wankhede Stadium            72
Rajiv Gandhi International Stadium, Uppal  63
M Chinnaswamy Stadium       62
MA Chidambaram Stadium, Chepauk  56
Sawai Mansingh Stadium       47
Punjab Cricket Association Stadium, Mohali  35
Sheikh Zayed Stadium         27
Dubai International Cricket Stadium  23
Name: count, dtype: int64
```

```
In [57]: ipl_nonna_df["venue"].value_counts().head(10).plot(kind="bar")
```

```
Out[57]: <Axes: xlabel='venue'>
```

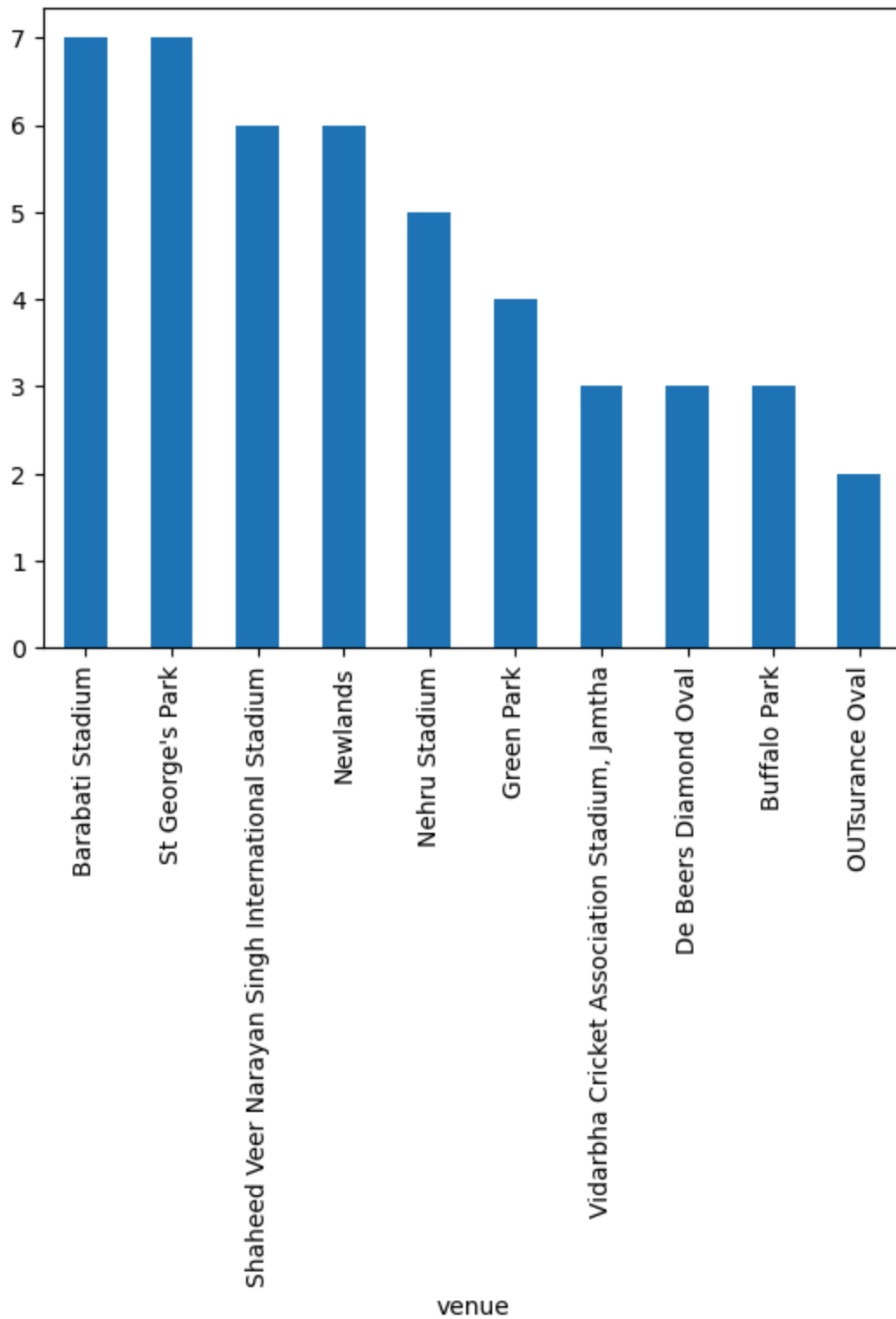


```
In [58]: ipl_nonna_df["venue"].value_counts().tail(10)
```

```
Out[58]: venue
Barabati Stadium 7
St George's Park 7
Shaheed Veer Narayan Singh International Stadium 6
Newlands 6
Nehru Stadium 5
Green Park 4
Vidarbha Cricket Association Stadium, Jamtha 3
De Beers Diamond Oval 3
Buffalo Park 3
OUTsurance Oval 2
Name: count, dtype: int64
```

```
In [59]: ipl_nonna_df["venue"].value_counts().tail(10).plot(kind = "bar")
```

```
Out[59]: <Axes: xlabel='venue'>
```



```
In [60]: ipl_nonna_df["team1"].nunique()
```

```
Out[60]: 15
```

```
In [61]: len(set(ipl_nonna_df["team1"]))
```

```
Out[61]: 15
```

```
In [62]: len(set(ipl_nonna_df["team1"].str.lower()))
```

```
Out[62]: 15
```

There is non duplicate data team1 column

```
In [64]: ipl_nonna_df["team1"].value_counts()
```

```
Out[64]: team1
Royal Challengers Bangalore    101
Mumbai Indians                 94
Chennai Super Kings            92
Kolkata Knight Riders          91
Kings XI Punjab               90
Delhi Daredevils               80
Rajasthan Royals               67
Sunrisers Hyderabad           56
Deccan Chargers                39
Pune Warriors                  23
Delhi Capitals                  17
Gujarat Lions                  15
Kochi Tuskers Kerala           7
Rising Pune Supergiants        7
Rising Pune Supergiant         7
Name: count, dtype: int64
```

```
In [65]: ipl_nonna_df["team1"].value_counts().sum()
```

```
Out[65]: 786
```

```
In [66]: ipl4 = ipl_nonna_df.replace({"Delhi Daredevils": "Delhi Capitals", "Rising Pune Super
```

```
In [67]: len(ipl4["team1"].value_counts())
```

```
Out[67]: 12
```

```
In [68]: ipl4["team1"].value_counts()
```

```
Out[68]: team1
Royal Challengers Bangalore    101
Delhi Capitals                  97
Sunrisers Hyderabad            95
Mumbai Indians                 94
Chennai Super Kings            92
Kolkata Knight Riders          91
Kings XI Punjab               90
Rajasthan Royals               67
Pune Warriors                  23
Gujarat Lions                  15
Rising Pune Supergiant         14
Kochi Tuskers Kerala           7
Name: count, dtype: int64
```

