In [ ]:
```
#2008-2019
```

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

In [5]:
```
ipl_df = pd.read_csv("matches.csv")
ipl_df.head()
```

Out[5]:

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

In [7]:
```
ipl_df.shape
```

Out[7]:
```
(756, 18)
```

In [8]:
```
ipl_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 756 entries, 0 to 755
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   id               756 non-null    int64 
 1   Season           756 non-null    object
 2   city             749 non-null    object
 3   date             756 non-null    object
 4   team1            756 non-null    object
 5   team2            756 non-null    object
 6   toss_winner      756 non-null    object
 7   toss_decision    756 non-null    object
 8   result           756 non-null    object
 9   dl_applied       756 non-null    int64 
 10  winner           752 non-null    object
 11  win_by_runs      756 non-null    int64 
 12  win_by_wickets   756 non-null    int64 
 13  player_of_match  752 non-null    object
 14  venue            756 non-null    object
 15  umpire1          754 non-null    object
 16  umpire2          754 non-null    object
 17  umpire3          119 non-null    object
dtypes: int64(4), object(14)
memory usage: 106.4+ KB
```

In [6]:
```
deliveries_df = pd.read_csv("deliveries.csv")
deliveries_df.head()
```

Out[6]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_s
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	

5 rows × 21 columns

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

Out[10]:
```
(179078, 21)
```

In [11]:
```
deliveries_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 179078 entries, 0 to 179077
Data columns (total 21 columns):
 #   Column            Non-Null Count   Dtype 
---  ------            --------------   ----- 
 0   match_id          179078 non-null  int64 
 1   inning            179078 non-null  int64 
 2   batting_team      179078 non-null  object
 3   bowling_team      179078 non-null  object
 4   over              179078 non-null  int64 
 5   ball              179078 non-null  int64 
 6   batsman           179078 non-null  object
 7   non_striker       179078 non-null  object
 8   bowler            179078 non-null  object
 9   is_super_over     179078 non-null  int64 
 10  wide_runs         179078 non-null  int64 
 11  bye_runs          179078 non-null  int64 
 12  legbye_runs       179078 non-null  int64 
 13  noball_runs       179078 non-null  int64 
 14  penalty_runs      179078 non-null  int64 
 15  batsman_runs      179078 non-null  int64 
 16  extra_runs        179078 non-null  int64 
 17  total_runs        179078 non-null  int64 
 18  player_dismissed  8834 non-null    object
 19  dismissal_kind    8834 non-null    object
 20  fielder           6448 non-null    object
dtypes: int64(13), object(8)
memory usage: 28.7+ MB
```

In [13]:
```
discard_columns = ["umpire1","umpire2","umpire3"]
ipl_df.drop(discard_columns,axis=1,inplace=True)
```

In [14]:
```
ipl_df.shape
```

Out[14]:
```
(756, 15)
```

In [18]:
```
ipl_df.result.value_counts()
```

Out[18]:
```
normal       743
tie            9
no result      4
Name: result, dtype: int64
```

Number of matches hosted in differenr cities

In [21]:
```
city_host = ipl_df.city.value_counts()
city_host.head()
```

Out[21]:
```
Mumbai       101
Kolkata       77
Delhi         74
Bangalore     66
Hyderabad     64
Name: city, dtype: int64
```

In [30]:
```
city_host.index
```

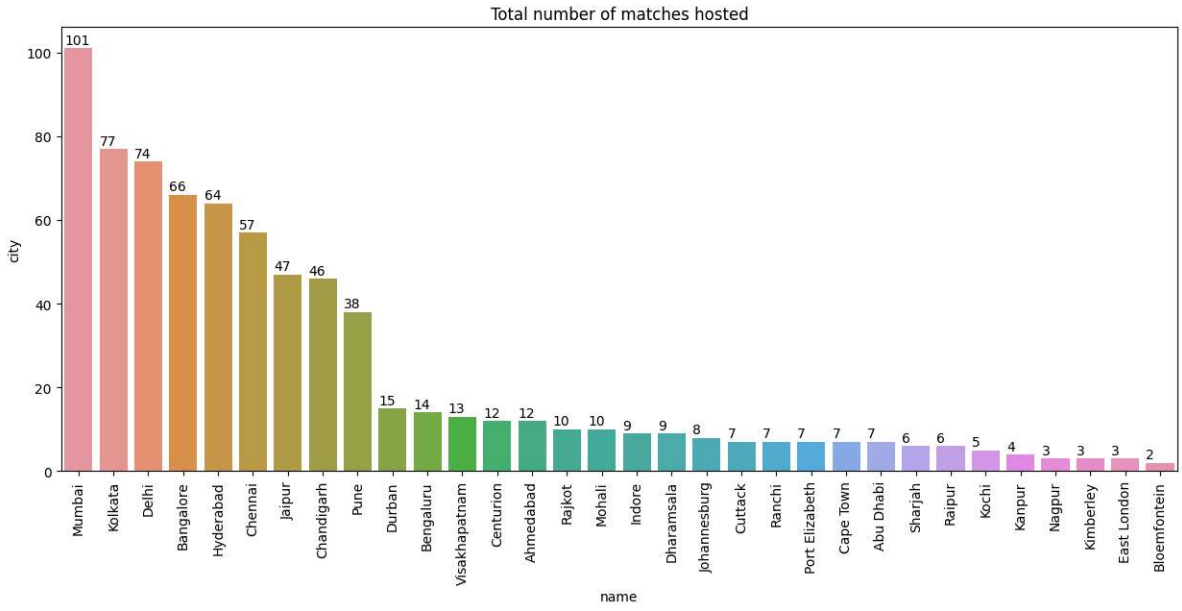
Out[30]:
```
Index(['Mumbai', 'Kolkata', 'Delhi', 'Bangalore', 'Hyderabad', 'Chennai',
       'Jaipur', 'Chandigarh', 'Pune', 'Durban', 'Bengaluru', 'Visakhapatna
m',
       'Centurion', 'Ahmedabad', 'Rajkot', 'Mohali', 'Indore', 'Dharamsala',
       'Johannesburg', 'Cuttack', 'Ranchi', 'Port Elizabeth', 'Cape Town',
       'Abu Dhabi', 'Sharjah', 'Raipur', 'Kochi', 'Kanpur', 'Nagpur',
       'Kimberley', 'East London', 'Bloemfontein'],
      dtype='object')
```

In [39]:
```
cities
```

Out[39]:

	city	name
Mumbai	101	Mumbai
Kolkata	77	Kolkata
Delhi	74	Delhi
Bangalore	66	Bangalore
Hyderabad	64	Hyderabad
Chennai	57	Chennai
Jaipur	47	Jaipur
Chandigarh	46	Chandigarh
Pune	38	Pune
Durban	15	Durban
Bengaluru	14	Bengaluru
Visakhapatnam	13	Visakhapatnam
Centurion	12	Centurion
Ahmedabad	12	Ahmedabad
Rajkot	10	Rajkot
Mohali	10	Mohali
Indore	9	Indore
Dharamsala	9	Dharamsala
Johannesburg	8	Johannesburg
Cuttack	7	Cuttack
Ranchi	7	Ranchi
Port Elizabeth	7	Port Elizabeth
Cape Town	7	Cape Town
Abu Dhabi	7	Abu Dhabi
Sharjah	6	Sharjah
Raipur	6	Raipur
Kochi	5	Kochi
Kanpur	4	Kanpur
Nagpur	3	Nagpur
Kimberley	3	Kimberley
East London	3	East London
Bloemfontein	2	Bloemfontein

In [68]:
```
plt.figure(figsize=(15,6))
plt.xticks(rotation = 90)
plt.title("Total number of matches hosted")
count = 0
cities = pd.DataFrame(city_host)
cities['name'] = cities.index
for i in cities['city']:
    plt.text(count- 0.4,i+1, str(i),color='black')
    count+=1
sns.barplot(cities, y= 'city',x='name' )
plt.show()
```



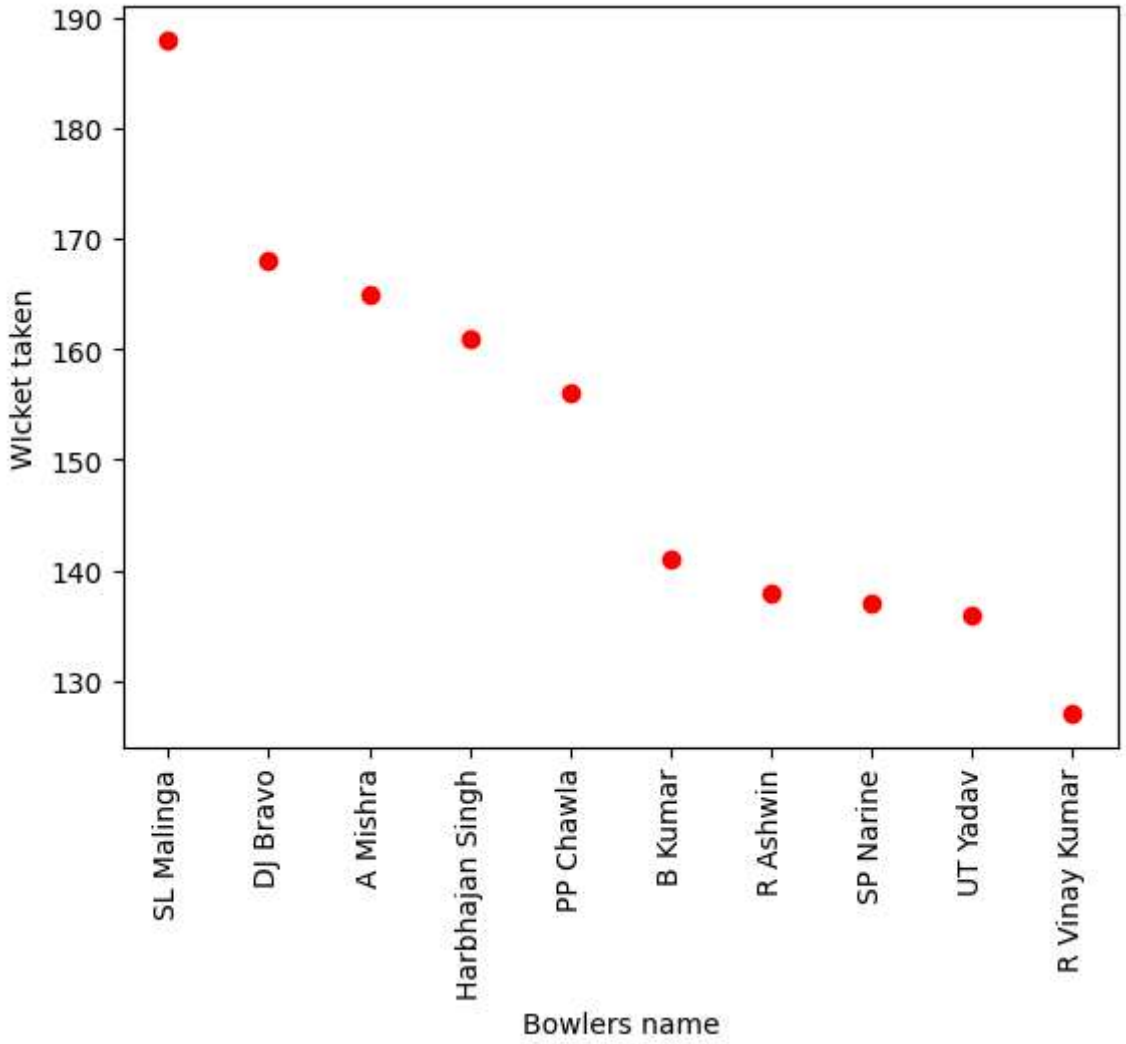
Maximum wickets taken by bowler in last 10 seasons

In [84]:
```
bowling_wickets = deliveries_df[deliveries_df["dismissal_kind"]!='run_out']
bowling_total = bowling_wickets.groupby('bowler').apply(lambda x : x["dismissa
bowling_wicket_count = bowling_total.groupby('bowler').count().reset_index()
bowling_top = bowling_wicket_count.sort_values(by='wickets',ascending=False)
top_bowlers = bowling_top.loc[:,["bowler","wickets"]][0:10]
top_bowlers
```

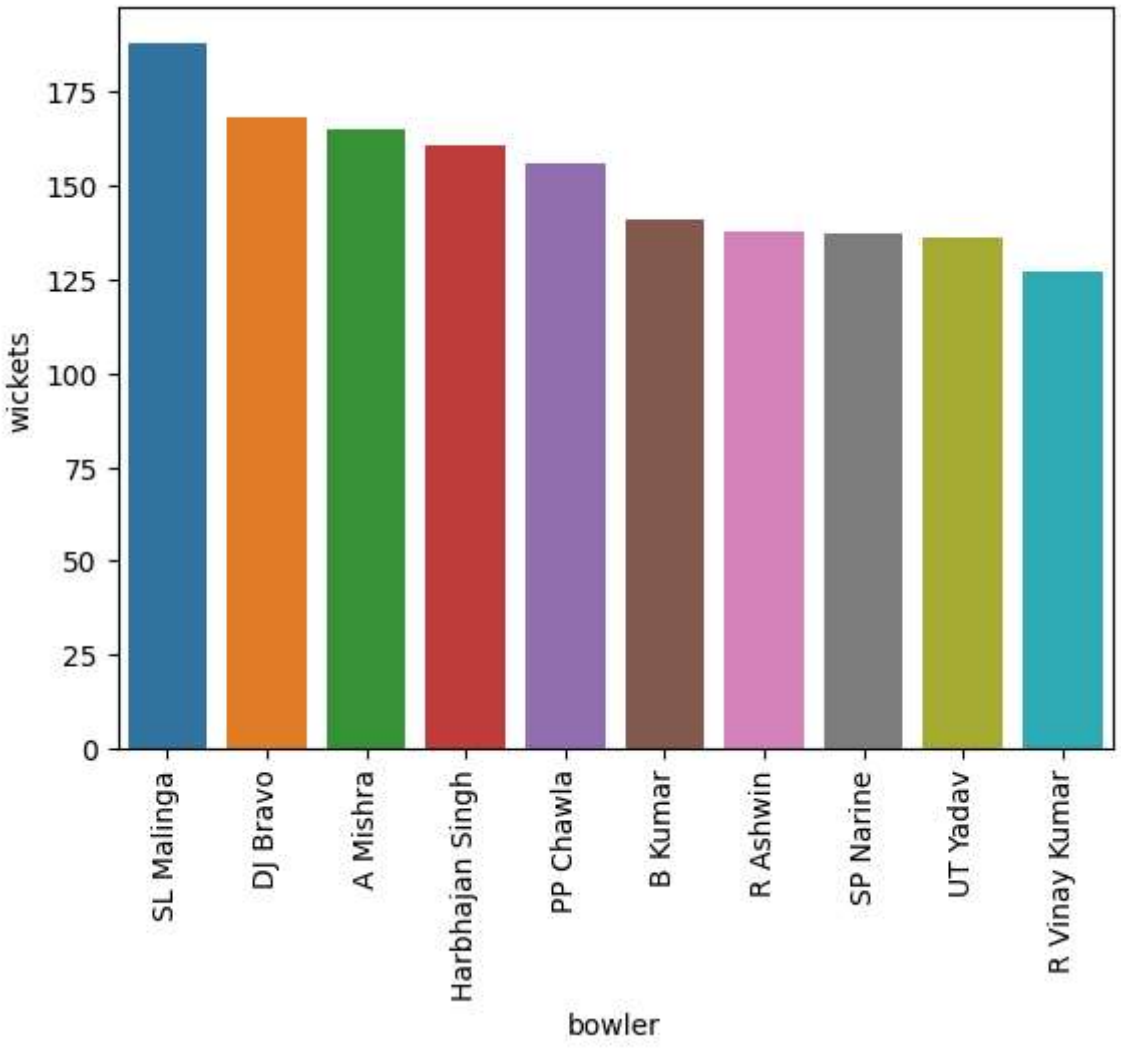
Out[84]:

	bowler	wickets
294	SL Malinga	188
84	DJ Bravo	168
6	A Mishra	165
113	Harbhajan Singh	161
231	PP Chawla	156
48	B Kumar	141
236	R Ashwin	138
300	SP Narine	137
330	UT Yadav	136
247	R Vinay Kumar	127

In [94]:
```
plt.scatter(top_bowlers["bowler"],top_bowlers["wickets"],color='r')
plt.xticks(rotation= 90)
plt.xlabel("Bowlers name")
plt.ylabel("WIcket taken")
plt.show()
```



In [91]:
```
sns.barplot(top_bowlers ,x = top_bowlers["bowler"],y = top_bowlers["wickets"])
plt.xticks(rotation= 90)
plt.show()
```



In [ ]:
```
```