In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:

```
df=pd.read_csv('IPL-DATA.csv')
df
```

Out[2]:

	id	city	date	player_of_match	venue	neutral_venue	tea
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	R Challen Banga
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	King Pu
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	D Dared
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mur Ind
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kol Kn Ri
...	...	...	...	...	...	...	
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0	R Challen Banga
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mur Ind
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0	R Challen Banga
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0	D Cap
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0	D Cap

816 rows × 17 columns

In [3]:

```
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 [4]:

```
duplicated_rows=df[df.duplicated()]
print("duplicate rows based on all columns:")
print(duplicated_rows)
```

```
duplicate rows based on all columns:
Empty DataFrame
Columns: [id, city, date, player_of_match, venue, neutral_venue, team1, team2, to
ss_winner, toss_decision, winner, result, result_margin, eliminator, method, umpi
re1, umpire2]
Index: []
```

In [5]:

```
total_sum_duplicated_valuse=df.duplicated().sum()
print(total_sum_duplicated_valuse)
```

```
0
```

In [6]:

```
df.columns
```

Out[6]:

```
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 [7]:

```
df.describe().T
```

Out[7]:

	count	mean	std	min	25%	50%	
id	816.0	756349.553922	305894.254804	335982.0	501227.75	729298.0	108
neutral_venue	816.0	0.094363	0.292512	0.0	0.00	0.0	
result_margin	799.0	17.321652	22.068427	1.0	6.00	8.0	

In [8]:

```
df.isnull().sum()
```

Out[8]:

```
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 [9]:

```
df.head(2)
```

Out[9]:

	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

In [10]:

```
temp=pd.DataFrame({"winner":df['winner']})
count=temp.value_counts()
print(count)
```

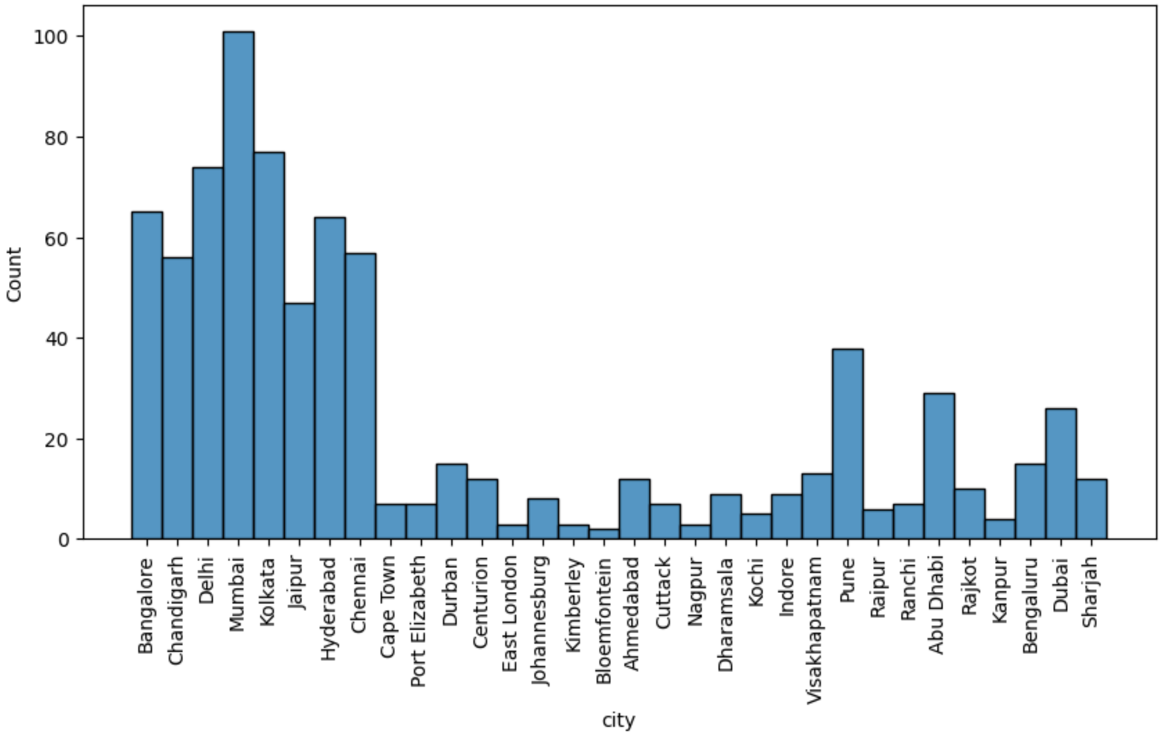
```
winner
Mumbai Indians                 120
Chennai Super Kings            106
Kolkata Knight Riders           99
Royal Challengers Bangalore     91
Kings XI Punjab                 88
Rajasthan Royals                81
Delhi Daredevils                67
Sunrisers Hyderabad             66
Deccan Chargers                 29
Delhi Capitals                  19
Gujarat Lions                   13
Pune Warriors                   12
Rising Pune Supergiant          10
Kochi Tuskers Kerala             6
Rising Pune Supergiants          5
Name: count, dtype: int64
```

In [11]:

```
plt.figure(figsize=(10,5))
sns.histplot(x='city',data=df,palette="Set2")
plt.xticks(rotation=90)
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_13672\1128625387.py:2: UserWarning: Igno
ring `palette` because no `hue` variable has been assigned.
  sns.histplot(x='city',data=df,palette="Set2")
```

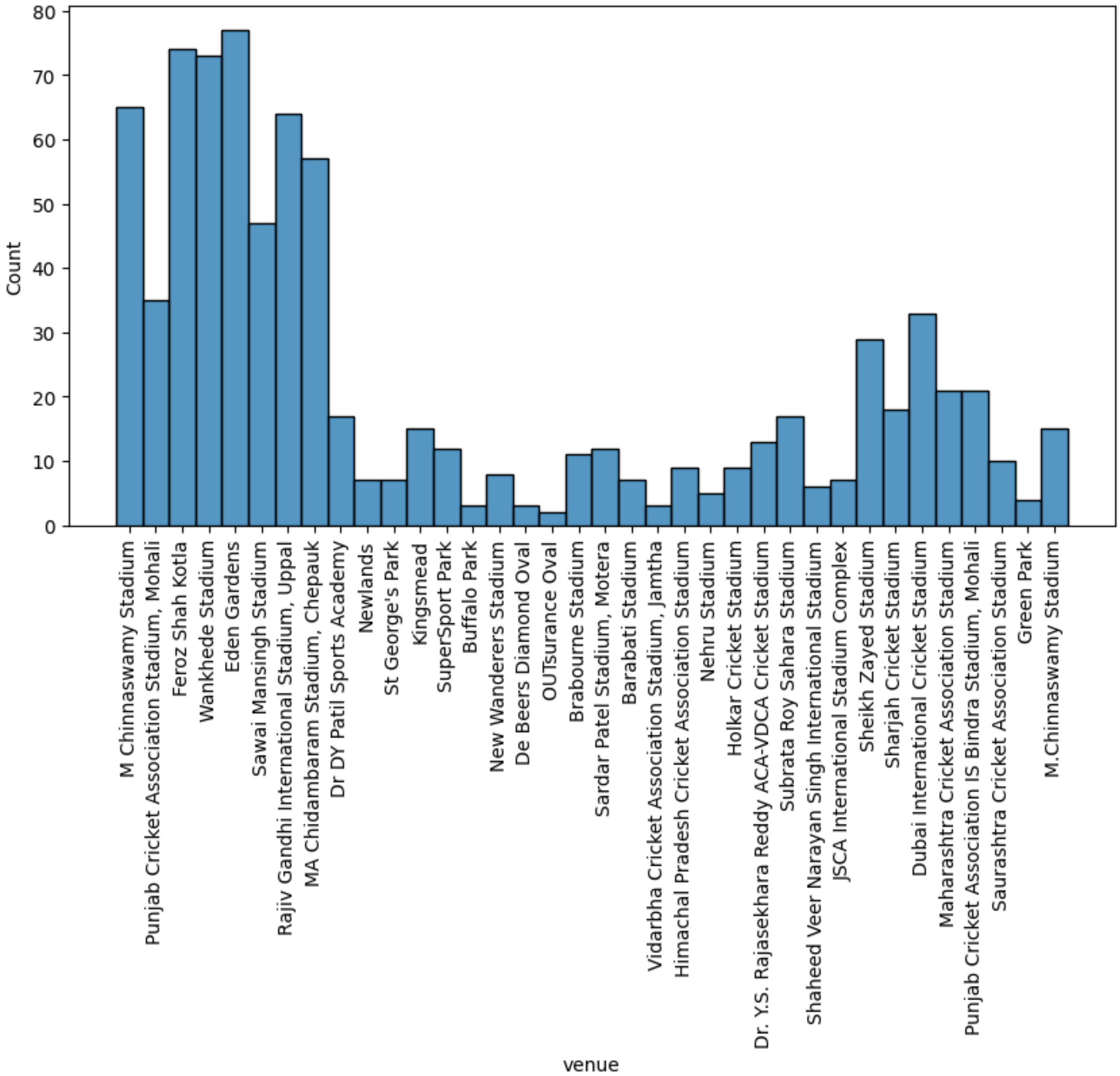
```
Out[11]: ([0,
           1,
           2,
           3,
           4,
           5,
           6,
           7,
           8,
           9,
           10,
           11,
           12,
           13,
           14,
           15,
           16,
           17,
           18,
           19,
           20,
           21,
           22,
           23,
           24,
           25,
           26,
           27,
           28,
           29,
           30,
           31],
          [Text(0, 0, 'Bangalore'),
           Text(1, 0, 'Chandigarh'),
           Text(2, 0, 'Delhi'),
           Text(3, 0, 'Mumbai'),
           Text(4, 0, 'Kolkata'),
           Text(5, 0, 'Jaipur'),
           Text(6, 0, 'Hyderabad'),
           Text(7, 0, 'Chennai'),
           Text(8, 0, 'Cape Town'),
           Text(9, 0, 'Port Elizabeth'),
           Text(10, 0, 'Durban'),
           Text(11, 0, 'Centurion'),
           Text(12, 0, 'East London'),
           Text(13, 0, 'Johannesburg'),
           Text(14, 0, 'Kimberley'),
           Text(15, 0, 'Bloemfontein'),
           Text(16, 0, 'Ahmedabad'),
           Text(17, 0, 'Cuttack'),
           Text(18, 0, 'Nagpur'),
           Text(19, 0, 'Dharamsala'),
           Text(20, 0, 'Kochi'),
           Text(21, 0, 'Indore'),
           Text(22, 0, 'Visakhapatnam'),
           Text(23, 0, 'Pune'),
           Text(24, 0, 'Raipur'),
           Text(25, 0, 'Ranchi'),
           Text(26, 0, 'Abu Dhabi'),
           Text(27, 0, 'Rajkot'),
           Text(28, 0, 'Kanpur'),
           Text(29, 0, 'Bengaluru'),
           Text(30, 0, 'Dubai'),
           Text(31, 0, 'Sharjah')])
```



In [12]:
```
plt.figure(figsize=(10,5))
sns.histplot(x='venue',data=df,palette='crest')
plt.xticks(rotation=90)
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_13672\2030287645.py:2: UserWarning: Igno
ring `palette` because no `hue` variable has been assigned.
  sns.histplot(x='venue',data=df,palette='crest')
```

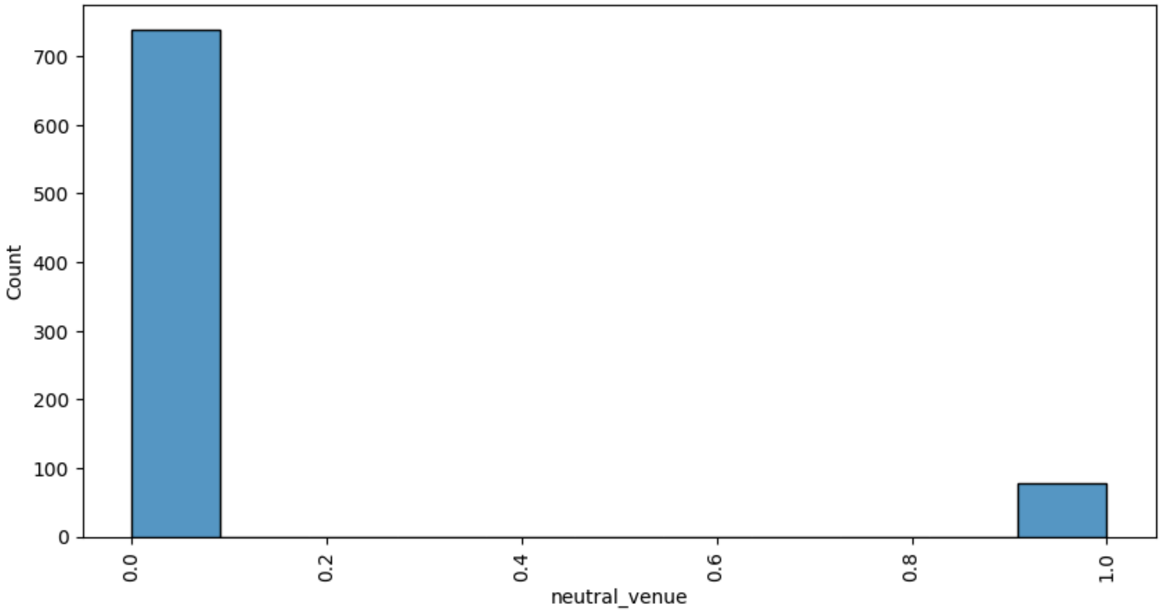
```
Out[12]:  ([0,
            1,
            2,
            3,
            4,
            5,
            6,
            7,
            8,
            9,
            10,
            11,
            12,
            13,
            14,
            15,
            16,
            17,
            18,
            19,
            20,
            21,
            22,
            23,
            24,
            25,
            26,
            27,
            28,
            29,
            30,
            31,
            32,
            33,
            34,
            35],
           [Text(0, 0, 'M Chinnaswamy Stadium'),
            Text(1, 0, 'Punjab Cricket Association Stadium, Mohali'),
            Text(2, 0, 'Feroz Shah Kotla'),
            Text(3, 0, 'Wankhede Stadium'),
            Text(4, 0, 'Eden Gardens'),
            Text(5, 0, 'Sawai Mansingh Stadium'),
            Text(6, 0, 'Rajiv Gandhi International Stadium, Uppal'),
            Text(7, 0, 'MA Chidambaram Stadium, Chepauk'),
            Text(8, 0, 'Dr DY Patil Sports Academy'),
            Text(9, 0, 'Newlands'),
            Text(10, 0, "St George's Park"),
            Text(11, 0, 'Kingsmead'),
            Text(12, 0, 'SuperSport Park'),
            Text(13, 0, 'Buffalo Park'),
            Text(14, 0, 'New Wanderers Stadium'),
            Text(15, 0, 'De Beers Diamond Oval'),
            Text(16, 0, 'OUTsurance Oval'),
            Text(17, 0, 'Brabourne Stadium'),
            Text(18, 0, 'Sardar Patel Stadium, Motera'),
            Text(19, 0, 'Barabati Stadium'),
            Text(20, 0, 'Vidarbha Cricket Association Stadium, Jamtha'),
            Text(21, 0, 'Himachal Pradesh Cricket Association Stadium'),
            Text(22, 0, 'Nehru Stadium'),
            Text(23, 0, 'Holkar Cricket Stadium'),
            Text(24, 0, 'Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium'),
            Text(25, 0, 'Subrata Roy Sahara Stadium'),
            Text(26, 0, 'Shaheed Veer Narayan Singh International Stadium'),
            Text(27, 0, 'JSCA International Stadium Complex'),
            Text(28, 0, 'Sheikh Zayed Stadium'),
            Text(29, 0, 'Sharjah Cricket Stadium'),
            Text(30, 0, 'Dubai International Cricket Stadium'),
            Text(31, 0, 'Maharashtra Cricket Association Stadium'),
            Text(32, 0, 'Punjab Cricket Association IS Bindra Stadium, Mohali'),
            Text(33, 0, 'Saurashtra Cricket Association Stadium'),
            Text(34, 0, 'Green Park'),
            Text(35, 0, 'M.Chinnaswamy Stadium')])
```



In [13]:
```
plt.figure(figsize=(10,5))
sns.histplot(x='neutral_venue',data=df,palette="Set2")
plt.xticks(rotation=90)
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_13672\3567544274.py:2: UserWarning: Igno
ring `palette` because no `hue` variable has been assigned.
  sns.histplot(x='neutral_venue',data=df,palette="Set2")
```

```
Out[13]: (array([-0.2,  0. ,  0.2,  0.4,  0.6,  0.8,  1. ,  1.2]),
          [Text(-0.2, 0, '−0.2'),
           Text(0.0, 0, '0.0'),
           Text(0.2, 0, '0.2'),
           Text(0.4000000000000001, 0, '0.4'),
           Text(0.6000000000000001, 0, '0.6'),
           Text(0.8, 0, '0.8'),
           Text(1.0000000000000002, 0, '1.0'),
           Text(1.2000000000000002, 0, '1.2')])
```

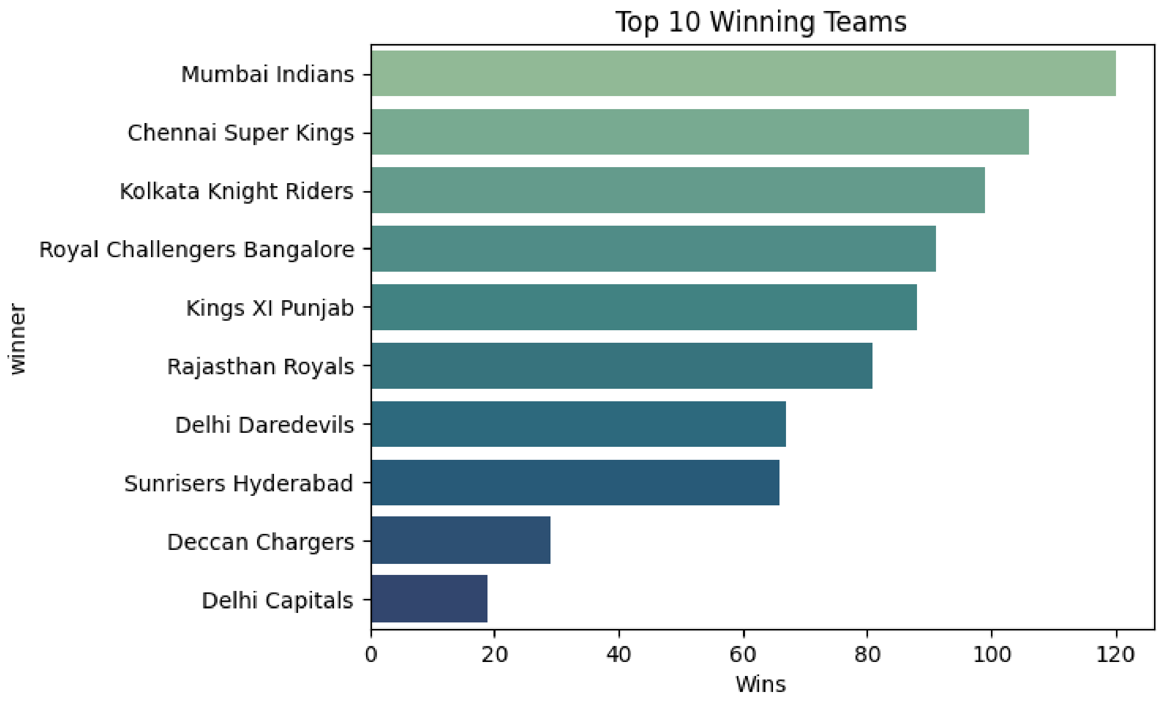


```
In [14]: total_matches = df.shape[0]
         print("Total matches:", total_matches)
```

```
Total matches: 816
```

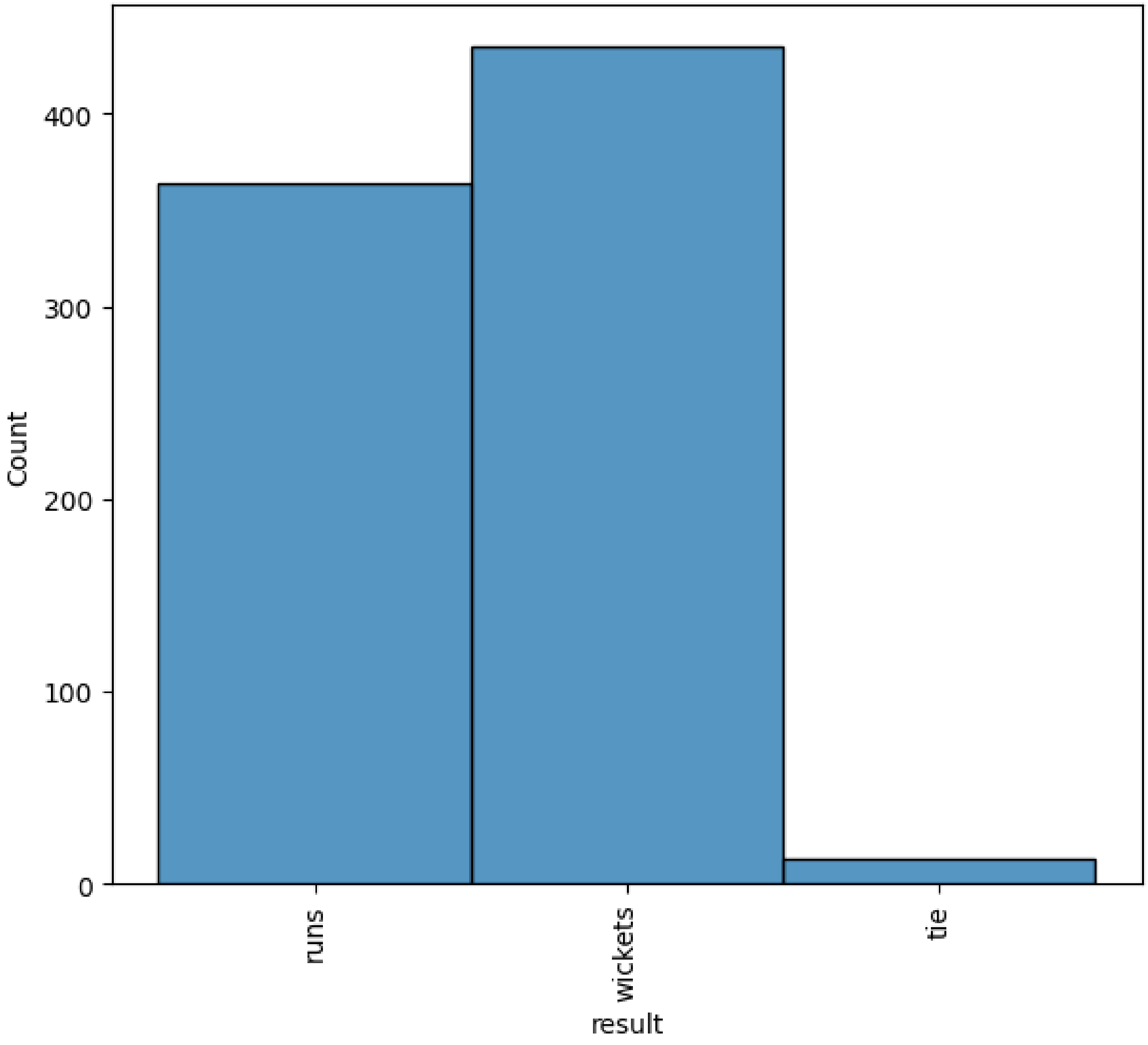
```
In [15]: top_teams = df['winner'].value_counts().head(10)
         sns.barplot(x=top_teams.values, y=top_teams.index, palette='crest')
         plt.title("Top 10 Winning Teams")
         plt.xlabel("Wins")
         plt.show()
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_13672\987869663.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v
0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effe
ct.

  sns.barplot(x=top_teams.values, y=top_teams.index, palette='crest')
```



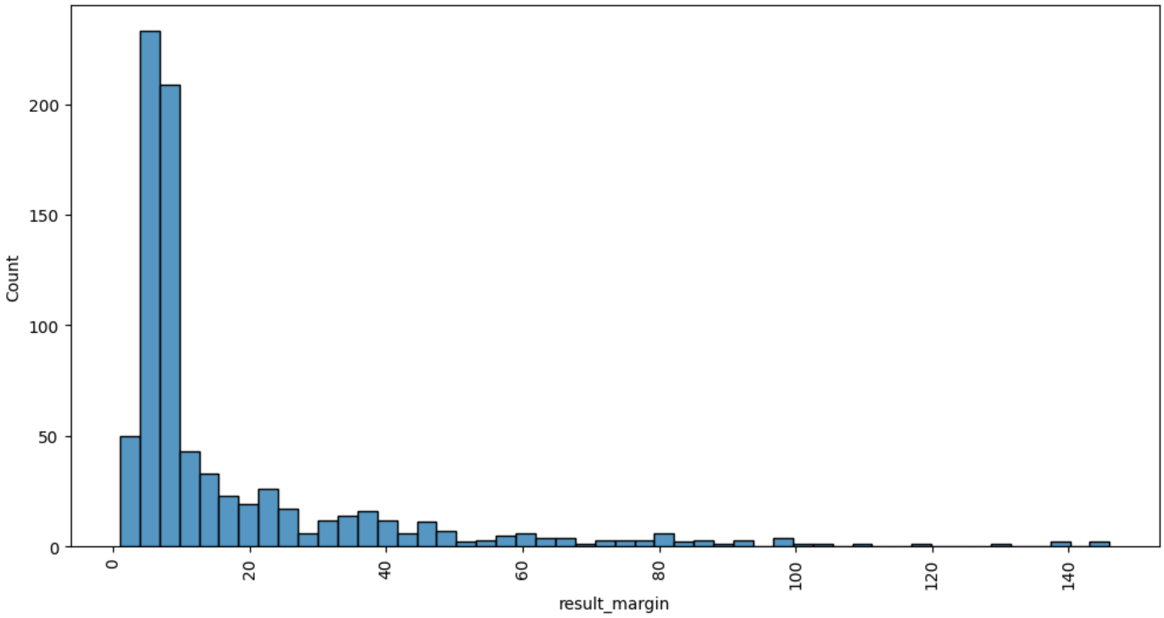
```
In [16]: plt.figure(figsize=(7,6))
         sns.histplot(x='result',data=df)
         plt.xticks(rotation=90)
```

```
Out[16]: ([0, 1, 2], [Text(0, 0, 'runs'), Text(1, 0, 'wickets'), Text(2, 0, 'tie')])
```



In [17]:
```
plt.figure(figsize=(12,6))
sns.histplot(x='result_margin',data=df)
plt.xticks(rotation=90)
```

Out[17]:
```
(array([-20.,   0.,  20.,  40.,  60.,  80., 100., 120., 140., 160.]),
 [Text(-20.0, 0, '-20'),
  Text(0.0, 0, '0'),
  Text(20.0, 0, '20'),
  Text(40.0, 0, '40'),
  Text(60.0, 0, '60'),
  Text(80.0, 0, '80'),
  Text(100.0, 0, '100'),
  Text(120.0, 0, '120'),
  Text(140.0, 0, '140'),
  Text(160.0, 0, '160')])
```



In [ ]:

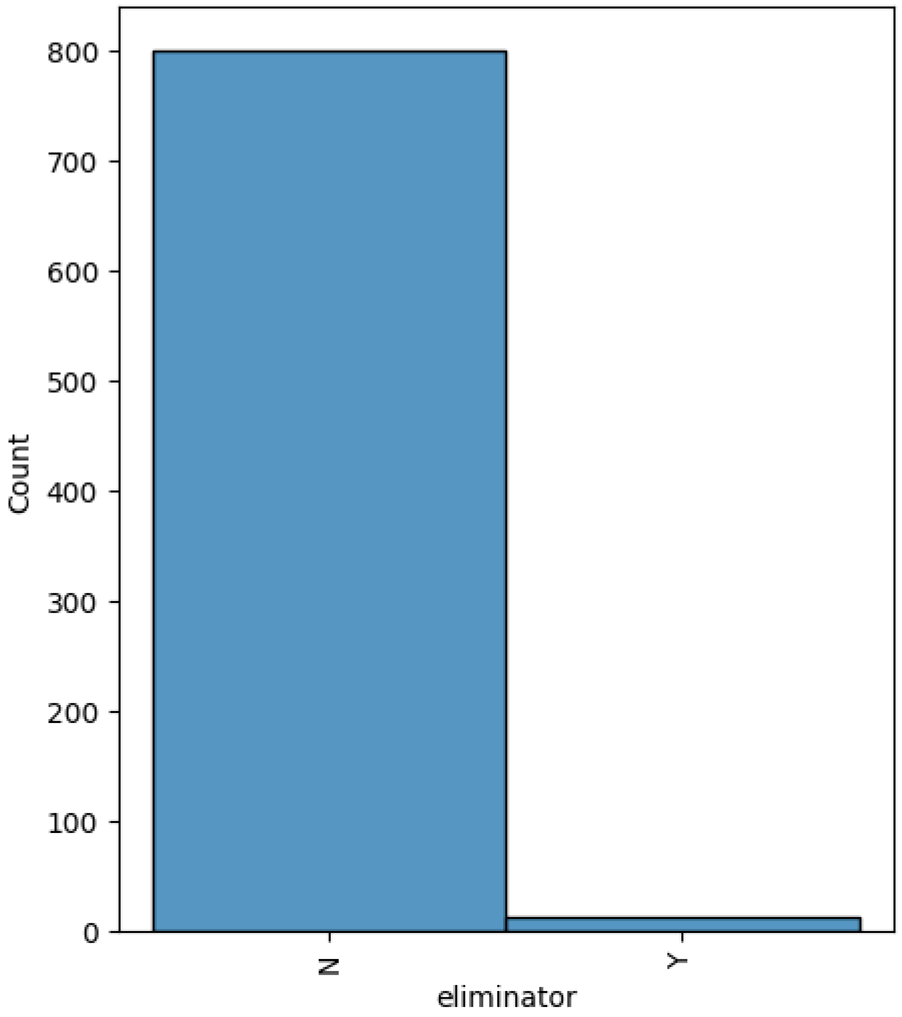
In [18]:
```
df.head(2)
```

Out[18]:

	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

In [19]:
```
plt.figure(figsize=(5,6))
sns.histplot(x='eliminator',data=df)
plt.xticks(rotation=90)
```

Out[19]:
```
([0, 1], [Text(0, 0, 'N'), Text(1, 0, 'Y')])
```



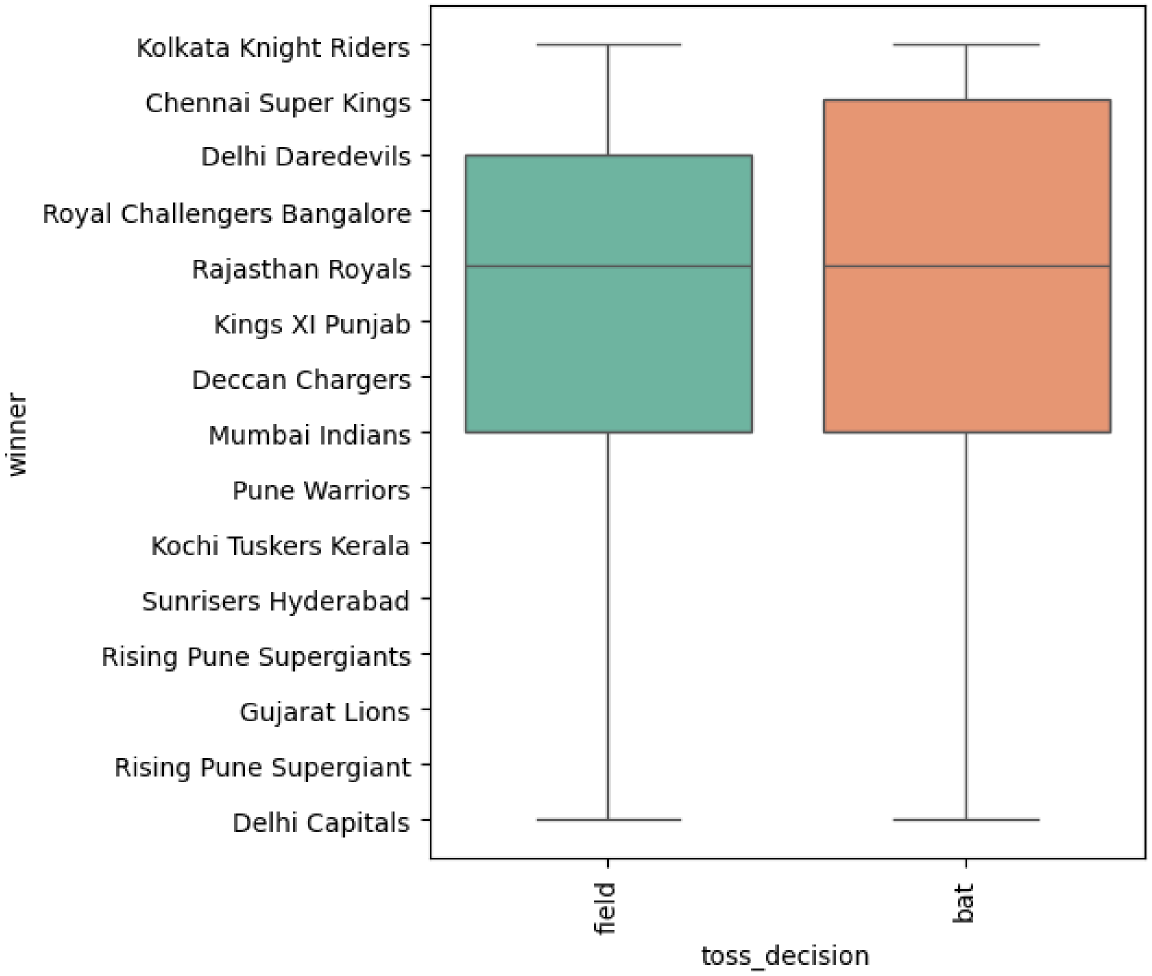
In [20]:

```
plt.figure(figsize=(5,6))
sns.boxplot(x='toss_decision',y='winner',data=df,palette="Set2")
plt.xticks(rotation=90)
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_13672\1032997117.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v
0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effe
ct.

  sns.boxplot(x='toss_decision',y='winner',data=df,palette="Set2")
```

Out[20]:

```
([0, 1], [Text(0, 0, 'field'), Text(1, 0, 'bat')])
```



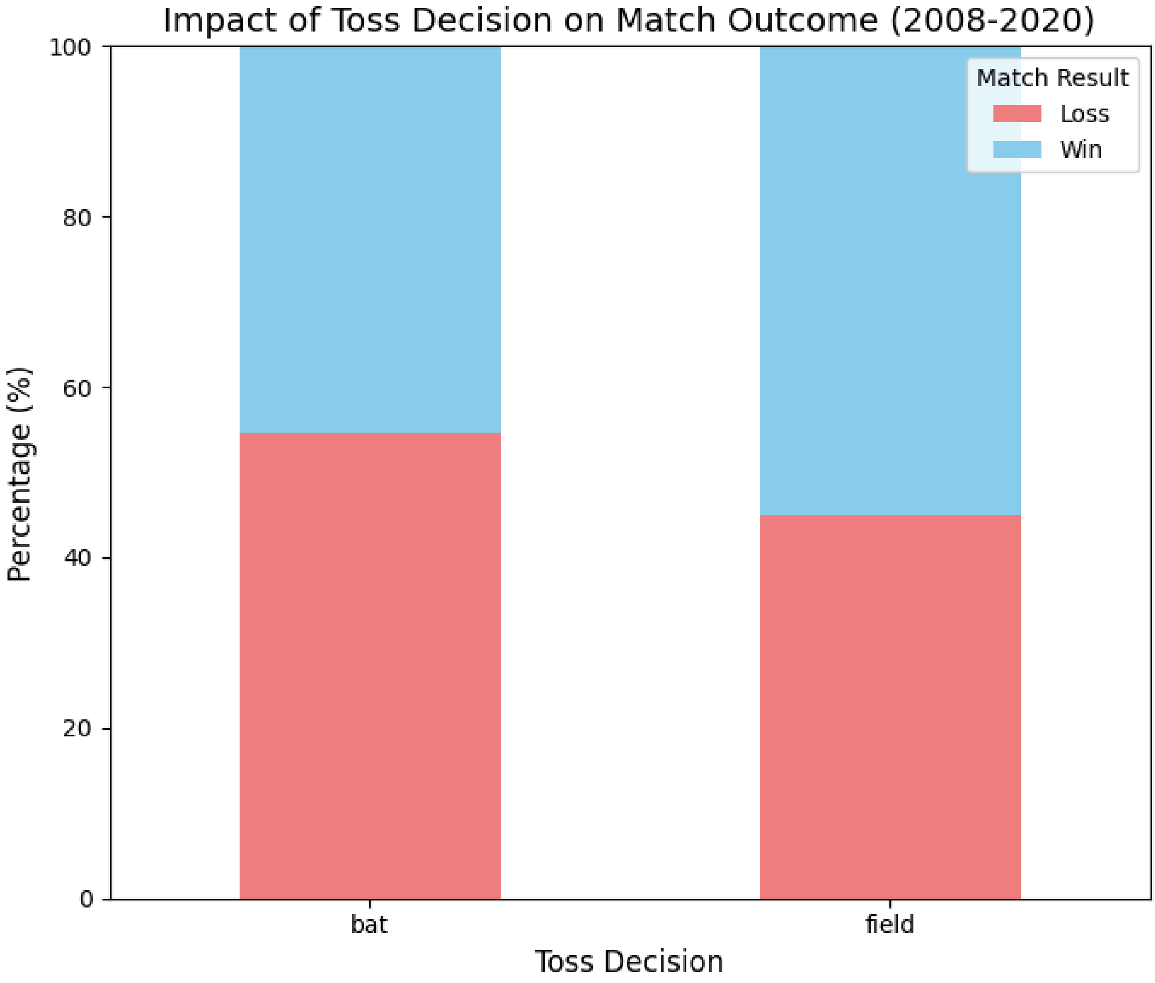
In [21]:
```
df.head(2)
```

Out[21]:

	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

In [22]:
```
df['Toss_Win_Match_Win'] = df.apply(
    lambda row: 'Win' if row['toss_winner'] == row['winner'] else 'Loss',
    axis=1
)
decision_vs_result = pd.crosstab(df['toss_decision'], df['Toss_Win_Match_Win'])
decision_vs_result_normalized = decision_vs_result.div(decision_vs_result.sum(ax
plt.figure(figsize=(7, 6))
decision_vs_result_normalized.plot(kind='bar', stacked=True, color=['lightcoral'

plt.title('Impact of Toss Decision on Match Outcome (2008-2020)', fontsize=14)
plt.xlabel('Toss Decision', fontsize=12)
plt.ylabel('Percentage (%)', fontsize=12)
plt.xticks(rotation=0)
plt.ylim(0, 100)
plt.legend(title='Match Result', loc='upper right')

plt.tight_layout()
plt.savefig("toss_decision_vs_match_outcome_final.png")
print("\n--- Toss Decision vs Match Outcome Percentage ---")
```

```
--- Toss Decision vs Match Outcome Percentage ---
```



```
In [26]: df['match_year'] = pd.to_datetime(df['date']).dt.year
         df['match_month'] = pd.to_datetime(df['date']).dt.month
```

Out[26]:

	id	city	date	player_of_match	venue	neutral_venue	tea
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	R Challen Banga
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	King Pu
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	D Dared
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mur Ind
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kol Kn Ri
...	...	...	...	...	...	...	
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0	R Challen Banga
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mur Ind
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0	R Challen Banga
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0	D Cap
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0	D Cap

816 rows × 20 columns

In [34]:

```
df['date'] = pd.to_datetime(df['date'], errors='coerce')
df.head(2)
```

Out[34]:

	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 rows × 21 columns

In [35]:

```
df['season'] = df['date'].dt.year
```

In [40]:

```
final_matches = df.groupby('season')['date'].max().reset_index()
final_matches.rename(columns={'date': 'final_date'}, inplace=True)
print("Final matches DataFrame:")
print(final_matches.head())
```

```
Final matches DataFrame:
   season final_date
0    2008 2008-06-01
1    2009 2009-05-24
2    2010 2010-04-25
3    2011 2011-05-28
4    2012 2012-05-27
```

In [46]:

```
df = pd.merge(df, final_matches, on='season', how='left')
print("\nColumns after merge:", df.columns)
df
```

```
Columns after merge: Index(['id', 'city', 'date', 'player_of_match', 'venue', 'ne
utral_venue',
       'team1', 'team2', 'toss_winner', 'toss_decision', 'winner', 'result',
       'result_margin', 'eliminator', 'method', 'umpire1', 'umpire2',
       'Toss_Win_Match_Win', 'match_year', 'match_month', 'season',
       'final_date_x', 'final_date_y', 'final_date'],
      dtype='object')
```

Out[46]:

	id	city	date	player_of_match	venue	neutral_venue	tea
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	R Challen Banga
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	King Pu
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	D Dared
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mur Ind
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kol Kn Ri
...	...	...	...	...	...	...	
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0	R Challen Banga
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mur Ind
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0	R Challen Banga
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0	D Cap
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0	D Cap

816 rows × 24 columns

In [48]:

```
def match_type(row):
    if row['eliminator'] == 'N':
        return 'League'
    elif row['eliminator'] == 'Y' and row['date'] == row['final_date']:
        return 'Final'
    elif row['eliminator'] == 'Y':
        return 'Eliminator'
    else:
        return 'Unknown'

df['match_type'] = df.apply(match_type, axis=1)
```

```
print("\nSample output:")
print(df[['season', 'date', 'eliminator', 'match_type']].head(15))
```

```
Sample output:
    season       date eliminator match_type
0     2008 2008-04-18          N     League
1     2008 2008-04-19          N     League
2     2008 2008-04-19          N     League
3     2008 2008-04-20          N     League
4     2008 2008-04-20          N     League
5     2008 2008-04-21          N     League
6     2008 2008-04-22          N     League
7     2008 2008-04-23          N     League
8     2008 2008-04-24          N     League
9     2008 2008-04-25          N     League
10    2008 2008-04-26          N     League
11    2008 2008-04-26          N     League
12    2008 2008-04-27          N     League
13    2008 2008-04-27          N     League
14    2008 2008-04-28          N     League
```

In [50]:

```
df
```

Out[50]:

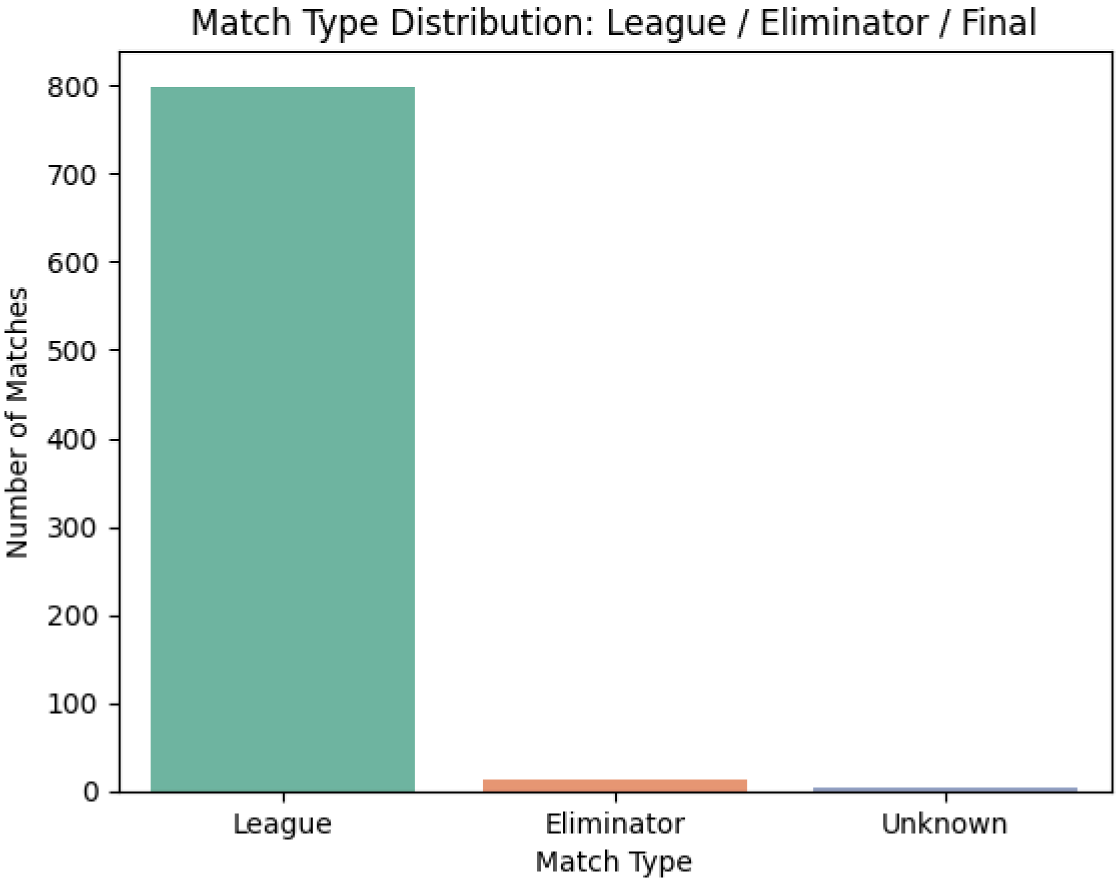
	id	city	date	player_of_match	venue	neutral_venue	tea
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	R Challen Banga
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	King Pu
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	D Dared
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mur Ind
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kol Kn Ri
...	...	...	...	...	...	...	
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0	R Challen Banga
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mur Ind
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0	R Challen Banga
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0	D Cap
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0	D Cap

816 rows × 25 columns

In [49]:

```
sns.countplot(x='match_type', data=df, palette='Set2')
plt.title('Match Type Distribution: League / Eliminator / Final')
plt.xlabel('Match Type')
plt.ylabel('Number of Matches')
plt.show()
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_13672\387320396.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v
0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effe
ct.

  sns.countplot(x='match_type', data=df, palette='Set2')
```



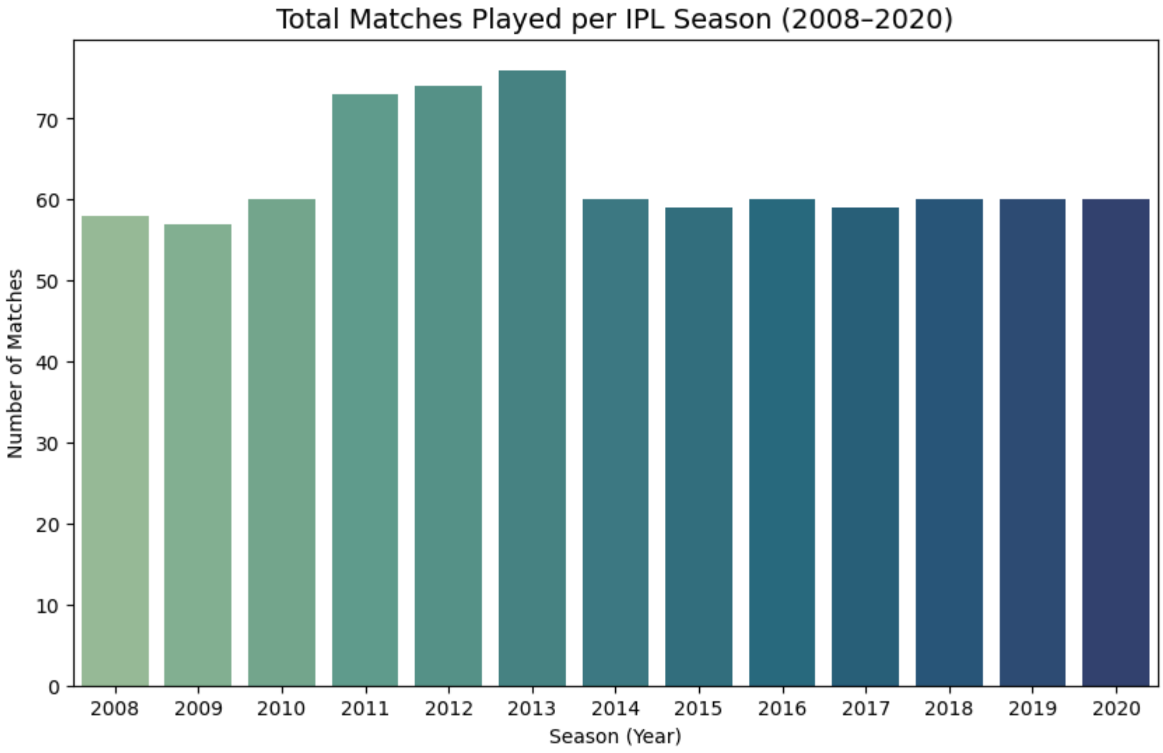
In [51]:
```
matches_per_season = df['season'].value_counts().sort_index()
```

In [52]:
```
print(matches_per_season)
```

```
season
2008    58
2009    57
2010    60
2011    73
2012    74
2013    76
2014    60
2015    59
2016    60
2017    59
2018    60
2019    60
2020    60
Name: count, dtype: int64
```

In [54]:
```
plt.figure(figsize=(10,6))
sns.barplot(x=matches_per_season.index, y=matches_per_season.values, palette='cr
plt.title('Total Matches Played per IPL Season (2008-2020)', fontsize=14)
plt.xlabel('Season (Year)')
plt.ylabel('Number of Matches')
plt.show()
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_13672\188864068.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v
0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effe
ct.

  sns.barplot(x=matches_per_season.index, y=matches_per_season.values, palette='c
rest')
```



```
In [ ]:
```