analysis

October 28, 2025

[1]:
```
#import libraries for analysis:
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

plt.style.use('dark_background')
plt.rcParams['axes.prop_cycle'] = plt.cycler(color = sns.color_palette('dark'))
plt.rcParams['figure.figsize'] = (20 , 20)
```

[2]:
```
# download dataset
#!curl -L -o ./data/ipl.csv "https://d2beiqkhq929f0.cloudfront.net/
↪public_assets/assets/000/158/850/original/IPL_2008-2024.csv?1760202529"
```

[ ]:
```
```

1 Question 1: Data Preprocessing & Feature Engineering (15 points)

1.0.1 Task:

1. Load the IPL dataset and perform comprehensive EDA
2. Handle missing values appropriately with justification
3. Create these new features:
    - `home_advantage`: Boolean indicating if team1 is playing in their home city
    - `match_importance`: Categorical (league/playoff/final) based on date and season
    - `toss_advantage`: Whether toss winner won the match
    - `season_phase`: Early/Mid/Late season

1.0.2 Deliverables:

- Clean dataset with no missing values
- Visualization showing distribution of matches across venues
- Statistical summary of win percentages for toss winners

[3]:
```
# load dataset
df = pd.read_csv('./data/ipl.csv')
df.head()
```

[3]:

```
       id  season        city        date match_type player_of_match  \
0  335982    2008   Bangalore  2008-04-18     League     BB McCullum
1  335983    2008  Chandigarh  2008-04-19     League      MEK Hussey
2  335984    2008       Delhi  2008-04-19     League     MF Maharoof
3  335985    2008      Mumbai  2008-04-20     League      MV Boucher
4  335986    2008     Kolkata  2008-04-20     League       DJ Hussey

                                        venue                        team1  \
0                       M Chinnaswamy Stadium  Royal Challengers Bangalore
1  Punjab Cricket Association Stadium, Mohali              Kings XI Punjab
2                            Feroz Shah Kotla             Delhi Daredevils
3                            Wankhede Stadium               Mumbai Indians
4                                Eden Gardens        Kolkata Knight Riders

                         team2                  toss_winner toss_decision  \
0        Kolkata Knight Riders  Royal Challengers Bangalore         field
1          Chennai Super Kings          Chennai Super Kings           bat
2             Rajasthan Royals             Rajasthan Royals           bat
3  Royal Challengers Bangalore               Mumbai Indians           bat
4              Deccan Chargers              Deccan Chargers           bat

                        winner   result  result_margin  target_runs  \
0        Kolkata Knight Riders     runs          140.0        223.0
1          Chennai Super Kings     runs           33.0        241.0
2             Delhi Daredevils  wickets            9.0        130.0
3  Royal Challengers Bangalore  wickets            5.0        166.0
4        Kolkata Knight Riders  wickets            5.0        111.0

  target_overs super_over method    umpire1         umpire2
0         20.0          N    NaN  Asad Rauf     RE Koertzen
1         20.0          N    NaN  MR Benson      SL Shastri
2         20.0          N    NaN  Aleem Dar  GA Pratapkumar
3         20.0          N    NaN   SJ Davis       DJ Harper
4         20.0          N    NaN  BF Bowden     K Hariharan
```

[4]:

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 20 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   id               1095 non-null   int64
 1   season           1095 non-null   int64
 2   city             1044 non-null   object
 3   date             1095 non-null   object
 4   match_type       1095 non-null   object
 5   player_of_match  1090 non-null   object
```

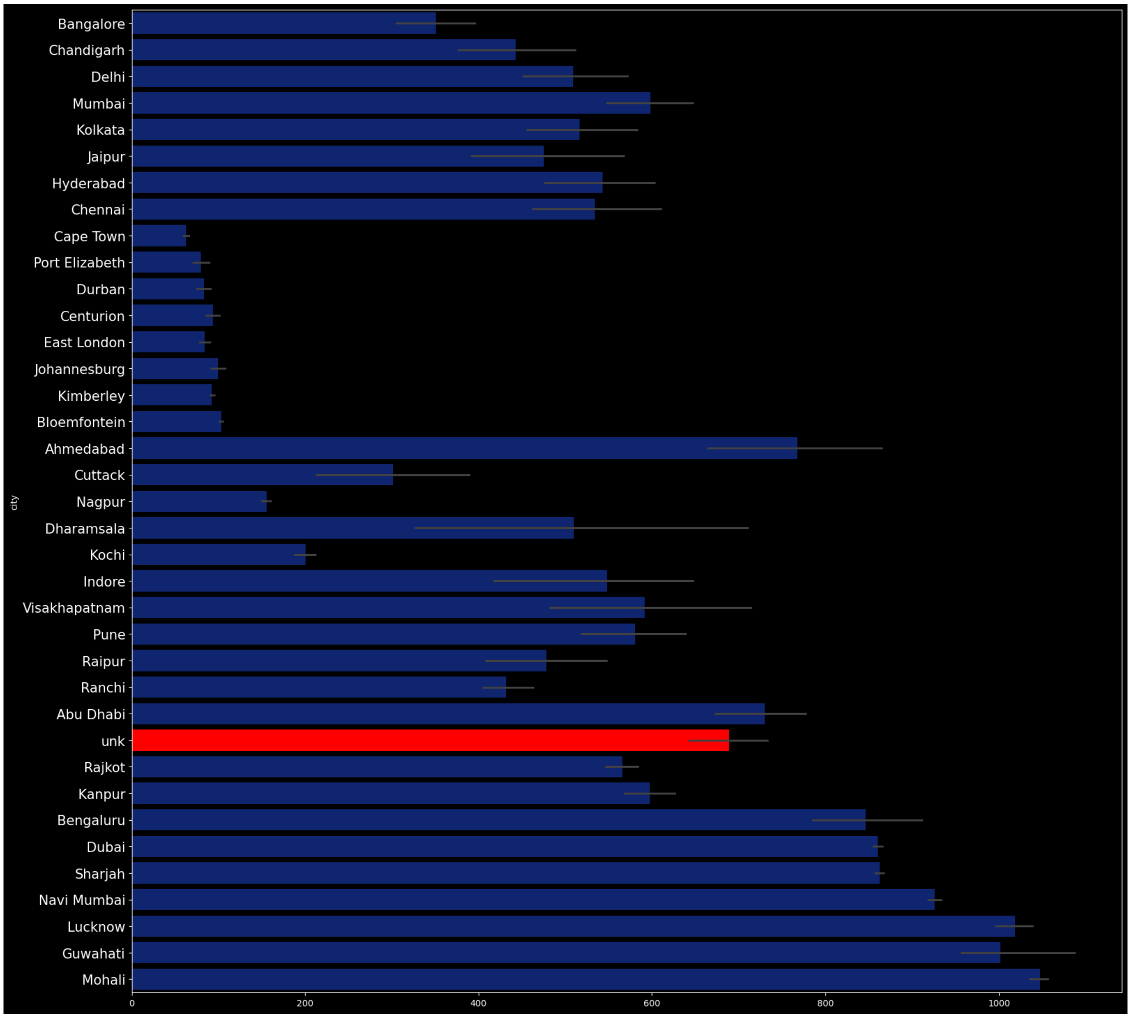
```
 6   venue            1095 non-null   object
 7   team1            1095 non-null   object
 8   team2            1095 non-null   object
 9   toss_winner      1095 non-null   object
 10  toss_decision    1095 non-null   object
 11  winner           1090 non-null   object
 12  result           1095 non-null   object
 13  result_margin    1076 non-null   float64
 14  target_runs      1092 non-null   float64
 15  target_overs     1092 non-null   float64
 16  super_over       1095 non-null   object
 17  method           21 non-null     object
 18  umpire1          1095 non-null   object
 19  umpire2          1095 non-null   object
dtypes: float64(3), int64(2), object(15)
memory usage: 171.2+ KB
```

1.1 EDA on indivisual columns:

1.2 1. city

[5]:
```python
# skip features such as id and session since they are not-null
# columns to check = ['city' , 'player_of_match' , 'winner' , 'result_margin' ,␣
↪'target_runs/overs' , 'method']

# 1. city:
city = df['city']

pmissing = (np.sum(city.isnull()) / city.size) * 100
print(f"Percentage of missing data: {np.round(pmissing , 2)}")

city = city.fillna('unk')
ax = sns.barplot(city)
bars = ax.patches
yticks = [t.get_text() for t in ax.get_yticklabels()]
yind = yticks.index('unk')
bars[yind].set_facecolor('red')

plt.yticks(fontsize = 15)
plt.show()

df['city'] = city

# stratergy -> replace missing data with unk token, to represent the set of␣
↪cities that hosted a match but were never mentioned.
```

```
Percentage of missing data: 4.66
```



1.2.1 2. player of match:

[6]:
```python
pom = df['player_of_match']
df[pom.isnull()]
```

[6]:
```
          id  season       city        date match_type player_of_match  \
241   501265    2011      Delhi  2011-05-21     League             NaN
485   829763    2015  Bangalore  2015-04-29     League             NaN
511   829813    2015  Bangalore  2015-05-17     League             NaN
744  1178424    2019  Bengaluru  2019-04-30     League             NaN
994  1359519    2023    Lucknow  2023-05-03     League             NaN

                                   venue  \
241                     Feroz Shah Kotla
485               M Chinnaswamy Stadium
511               M Chinnaswamy Stadium
```

```
744                             M.Chinnaswamy Stadium
994  Bharat Ratna Shri Atal Bihari Vajpayee Ekana C…

                           team1                team2  \
241             Delhi Daredevils        Pune Warriors
485  Royal Challengers Bangalore     Rajasthan Royals
511  Royal Challengers Bangalore     Delhi Daredevils
744  Royal Challengers Bangalore     Rajasthan Royals
994         Lucknow Super Giants  Chennai Super Kings

                     toss_winner toss_decision winner     result  \
241             Delhi Daredevils           bat    NaN  no result
485             Rajasthan Royals         field    NaN  no result
511  Royal Challengers Bangalore         field    NaN  no result
744             Rajasthan Royals         field    NaN  no result
994          Chennai Super Kings         field    NaN  no result

     result_margin  target_runs  target_overs super_over method  \
241            NaN          NaN           NaN          N    NaN
485            NaN          NaN           NaN          N    NaN
511            NaN        188.0          20.0          N    NaN
744            NaN         63.0           5.0          N    NaN
994            NaN          NaN           NaN          N    NaN

              umpire1        umpire2
241         SS Hazare      RJ Tucker
485         JD Cloete      PG Pathak
511  HDPK Dharmasena   K Srinivasan
744          NJ Llong      UV Gandhe
994      AK Chaudhary  NA Patwardhan
```

- since there are just 5 matches where no result was declared so both player_of_matcha and winner features are missing, we again replace the null value with unk token for sanity purposes(I love nlp!).

[7]:
```
df['player_of_match'] = df['player_of_match'].fillna('unk')
df['winner'] = df['winner'].fillna('unk')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 20 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   id               1095 non-null   int64
 1   season           1095 non-null   int64
 2   city             1095 non-null   object
 3   date             1095 non-null   object
```

```
 4   match_type       1095 non-null   object
 5   player_of_match  1095 non-null   object
 6   venue            1095 non-null   object
 7   team1            1095 non-null   object
 8   team2            1095 non-null   object
 9   toss_winner      1095 non-null   object
 10  toss_decision    1095 non-null   object
 11  winner           1095 non-null   object
 12  result           1095 non-null   object
 13  result_margin    1076 non-null   float64
 14  target_runs      1092 non-null   float64
 15  target_overs     1092 non-null   float64
 16  super_over       1095 non-null   object
 17  method           21 non-null     object
 18  umpire1          1095 non-null   object
 19  umpire2          1095 non-null   object
dtypes: float64(3), int64(2), object(15)
memory usage: 171.2+ KB
```

1.2.2 3. result_margin

[8]:
```
df[df['result_margin'].isnull()]
```

[8]:
```
          id  season       city        date match_type player_of_match  \
66    392190    2009  Cape Town  2009-04-23     League        YK Pathan
130   419121    2010    Chennai  2010-03-21     League        J Theron
241   501265    2011      Delhi  2011-05-21     League             unk
328   598004    2013  Hyderabad  2013-04-07     League       GH Vihari
342   598017    2013  Bangalore  2013-04-16     League         V Kohli
416   729315    2014  Abu Dhabi  2014-04-29     League     JP Faulkner
475   829741    2015  Ahmedabad  2015-04-21     League        SE Marsh
485   829763    2015  Bangalore  2015-04-29     League             unk
511   829813    2015  Bangalore  2015-05-17     League             unk
610  1082625    2017     Rajkot  2017-04-29     League       KH Pandya
705  1175365    2019      Delhi  2019-03-30     League         PP Shaw
744  1178424    2019  Bengaluru  2019-04-30     League             unk
746  1178426    2019     Mumbai  2019-05-02     League       JJ Bumrah
757  1216493    2020        unk  2020-09-20     League      MP Stoinis
765  1216547    2020        unk  2020-09-28     League  AB de Villiers
790  1216512    2020  Abu Dhabi  2020-10-18     League     LH Ferguson
791  1216517    2020        unk  2020-10-18     League        KL Rahul
835  1254077    2021    Chennai  2021-04-25     League         PP Shaw
994  1359519    2023    Lucknow  2023-05-03     League             unk

                                venue  \
66                           Newlands
130     MA Chidambaram Stadium, Chepauk
```

```
241                                        Feroz Shah Kotla
328               Rajiv Gandhi International Stadium, Uppal
342                                   M Chinnaswamy Stadium
416                                    Sheikh Zayed Stadium
475                            Sardar Patel Stadium, Motera
485                                   M Chinnaswamy Stadium
511                                   M Chinnaswamy Stadium
610                  Saurashtra Cricket Association Stadium
705                                    Arun Jaitley Stadium
744                                    M.Chinnaswamy Stadium
746                                        Wankhede Stadium
757                     Dubai International Cricket Stadium
765                     Dubai International Cricket Stadium
790                                    Sheikh Zayed Stadium
791                     Dubai International Cricket Stadium
835                MA Chidambaram Stadium, Chepauk, Chennai
994  Bharat Ratna Shri Atal Bihari Vajpayee Ekana C…

                           team1                        team2  \
66         Kolkata Knight Riders             Rajasthan Royals
130          Chennai Super Kings              Kings XI Punjab
241             Delhi Daredevils                Pune Warriors
328          Sunrisers Hyderabad  Royal Challengers Bangalore
342  Royal Challengers Bangalore             Delhi Daredevils
416        Kolkata Knight Riders             Rajasthan Royals
475             Rajasthan Royals              Kings XI Punjab
485  Royal Challengers Bangalore             Rajasthan Royals
511  Royal Challengers Bangalore             Delhi Daredevils
610                Gujarat Lions               Mumbai Indians
705        Kolkata Knight Riders               Delhi Capitals
744  Royal Challengers Bangalore             Rajasthan Royals
746               Mumbai Indians          Sunrisers Hyderabad
757               Delhi Capitals              Kings XI Punjab
765  Royal Challengers Bangalore               Mumbai Indians
790        Kolkata Knight Riders          Sunrisers Hyderabad
791               Mumbai Indians              Kings XI Punjab
835               Delhi Capitals          Sunrisers Hyderabad
994         Lucknow Super Giants          Chennai Super Kings

                     toss_winner toss_decision                       winner  \
66         Kolkata Knight Riders         field             Rajasthan Royals
130          Chennai Super Kings         field              Kings XI Punjab
241             Delhi Daredevils           bat                          unk
328  Royal Challengers Bangalore           bat          Sunrisers Hyderabad
342  Royal Challengers Bangalore         field  Royal Challengers Bangalore
416             Rajasthan Royals           bat             Rajasthan Royals
475              Kings XI Punjab         field              Kings XI Punjab
```

485	Rajasthan Royals	field	unk
511	Royal Challengers Bangalore	field	unk
610	Gujarat Lions	bat	Mumbai Indians
705	Delhi Capitals	field	Delhi Capitals
744	Rajasthan Royals	field	unk
746	Mumbai Indians	bat	Mumbai Indians
757	Kings XI Punjab	field	Delhi Capitals
765	Mumbai Indians	field	Royal Challengers Bangalore
790	Sunrisers Hyderabad	field	Kolkata Knight Riders
791	Mumbai Indians	bat	Kings XI Punjab
835	Delhi Capitals	bat	Delhi Capitals
994	Chennai Super Kings	field	unk

	result	result_margin	target_runs	target_overs	super_over	method \
66	tie	NaN	151.0	20.0	Y	NaN
130	tie	NaN	137.0	20.0	Y	NaN
241	no result	NaN	NaN	NaN	N	NaN
328	tie	NaN	131.0	20.0	Y	NaN
342	tie	NaN	153.0	20.0	Y	NaN
416	tie	NaN	153.0	20.0	Y	NaN
475	tie	NaN	192.0	20.0	Y	NaN
485	no result	NaN	NaN	NaN	N	NaN
511	no result	NaN	188.0	20.0	N	NaN
610	tie	NaN	154.0	20.0	Y	NaN
705	tie	NaN	186.0	20.0	Y	NaN
744	no result	NaN	63.0	5.0	N	NaN
746	tie	NaN	163.0	20.0	Y	NaN
757	tie	NaN	158.0	20.0	Y	NaN
765	tie	NaN	202.0	20.0	Y	NaN
790	tie	NaN	164.0	20.0	Y	NaN
791	tie	NaN	177.0	20.0	Y	NaN
835	tie	NaN	160.0	20.0	Y	NaN
994	no result	NaN	NaN	NaN	N	NaN

	umpire1	umpire2
66	MR Benson	M Erasmus
130	K Hariharan	DJ Harper
241	SS Hazare	RJ Tucker
328	AK Chaudhary	S Ravi
342	M Erasmus	VA Kulkarni
416	Aleem Dar	AK Chaudhary
475	M Erasmus	S Ravi
485	JD Cloete	PG Pathak
511	HDPK Dharmasena	K Srinivasan
610	AK Chaudhary	CB Gaffaney
705	AY Dandekar	Nitin Menon
744	NJ Llong	UV Gandhe

```
746        CK Nandan                 S Ravi
757      AK Chaudhary           Nitin Menon
765       Nitin Menon            PR Reiffel
790         PG Pathak                S Ravi
791       Nitin Menon            PR Reiffel
835       CB Gaffaney  KN Ananthapadmanabhan
994      AK Chaudhary         NA Patwardhan
```

1.2.3 for missing result_margin, its either a tie or no-result in match/ match abandoned. we replace NaN values with 0, since 0 margin = no-result/tie

[9]:
```python
df['result_margin'] = df['result_margin'].fillna(0)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 20 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   id               1095 non-null   int64
 1   season           1095 non-null   int64
 2   city             1095 non-null   object
 3   date             1095 non-null   object
 4   match_type       1095 non-null   object
 5   player_of_match  1095 non-null   object
 6   venue            1095 non-null   object
 7   team1            1095 non-null   object
 8   team2            1095 non-null   object
 9   toss_winner      1095 non-null   object
 10  toss_decision    1095 non-null   object
 11  winner           1095 non-null   object
 12  result           1095 non-null   object
 13  result_margin    1095 non-null   float64
 14  target_runs      1092 non-null   float64
 15  target_overs     1092 non-null   float64
 16  super_over       1095 non-null   object
 17  method           21 non-null     object
 18  umpire1          1095 non-null   object
 19  umpire2          1095 non-null   object
dtypes: float64(3), int64(2), object(15)
memory usage: 171.2+ KB
```

1.2.4 4. target_runs/overs:

[10]:
```python
truns = df['target_runs']
tovers = df['target_overs']
df[tovers.isnull()]
```

[10]:
```
          id  season       city        date match_type player_of_match  \
241   501265    2011      Delhi  2011-05-21     League             unk
485   829763    2015  Bangalore  2015-04-29     League             unk
994  1359519    2023    Lucknow  2023-05-03     League             unk

                                                venue  \
241                                  Feroz Shah Kotla
485                             M Chinnaswamy Stadium
994  Bharat Ratna Shri Atal Bihari Vajpayee Ekana C…

                           team1                team2          toss_winner  \
241              Delhi Daredevils        Pune Warriors     Delhi Daredevils
485  Royal Challengers Bangalore     Rajasthan Royals     Rajasthan Royals
994         Lucknow Super Giants  Chennai Super Kings  Chennai Super Kings

    toss_decision winner     result  result_margin  target_runs  target_overs  \
241           bat    unk  no result            0.0          NaN           NaN
485         field    unk  no result            0.0          NaN           NaN
994         field    unk  no result            0.0          NaN           NaN

    super_over method        umpire1        umpire2
241          N    NaN      SS Hazare      RJ Tucker
485          N    NaN      JD Cloete      PG Pathak
994          N    NaN  AK Chaudhary  NA Patwardhan
```

- https://www.espncricinfo.com/series/indian-premier-league-2011-466304/delhi-daredevils-vs-pune-warriors-68th-match-501265/full-scorecard
- I researched the first match and found out the game was abandoned due to extreme rain.
- This jibes with our previous pre-processing that says no-result for result feature.

1.2.5 we replace missing target_overs/runs values with zero, via the analogy no item = 0 item.

[11]:
```
df['target_runs'] = truns.fillna(0)
df['target_overs'] = tovers.fillna(0)
```

1.2.6 5. method feature:

- if a match finished under normal conditions then no external method to determine the winner of the match was used -> match finished under normal conditions - 1074/1095 matches.

[12]:
```
method = df['method']
print(f"% of matches finished normally: { np.sum(method.isnull()) / df.
↪shape[0]}")
df['method'] = method.fillna('normal')
```

```
% of matches finished normally: 0.9808219178082191
```

Q. Create these new features:

- `home_advantage`: Boolean indicating if team1 is playing in their home city
- `match_importance`: Categorical (league/playoff/final) based on date and season
- `toss_advantage`: Whether toss winner won the match
- `season_phase`: Early/Mid/Late season

[13]:
```python
# addressing the problem of teams that renamed themselves over the years
# 1. Royal Challengers Bangaluru -> Royal Challengers Bangalore
# 2. Kings XI Punjab -> Punjab Kings
# 3. Rising Pune Supergiants -> Rising Pune Supergiant
# 4. Delhi Daredevils -> Delhi Capitals
def change_team_name(x):
    if(x == 'Royal Challengers Bengaluru'):
        return 'Royal Challengers Bangalore'
    if(x == 'Kings XI Punjab'):
        return 'Punjab Kings'
    if(x == 'Rising Pune Supergiants'):
        return 'Rising Pune Supergiant'
    if(x == 'Delhi Daredevils'):
        return 'Delhi Capitals'
    return x
t1 = df['team1'].apply(lambda x: change_team_name(x))
t2 = df['team2'].apply(lambda x : change_team_name(x))
df['winner'] = df['winner'].apply(change_team_name)
df['team1'] = t1
df['team2'] = t2


print(f"cities: {df['city'].unique()} \n teams: {df['team1'].unique()}")

# create a map of team and home city
# team to city
ttc = {
    'Royal Challengers Bangalore' : ['Bangalore'],
    'Punjab Kings' : ["Mohali" , 'Chandigarh'],
    'Mumbai Indians' : ['Navi Mumbai' , 'Mumbai'],
    'Kolkata Knight Riders' : ['Kolkata'],
    'Rajasthan Royals' : ['Jaipur'],
    'Deccan Chargers' : ['Hyderabad' , 'Cuttuck' , 'Navi Mumbai'],      
 ↪#deccan had alternating home-grounds
    'Chennai Super Kings' : ['Chennai'],
    'Kochi Tuskers Kerala' : ['Kochi'],
    'Pune Warriors' : ['Pune'],
    'Sunrisers Hyderabad' : ['Hyderabad'],
    'Gujarat Lions': ['Rajkot'],
    'Rising Pune Supergiant' : ['Pune'],
    'Delhi Capitals' : ["Delhi"],
    'Lucknow Super Giants' : ['Lucknow'],
```

```python
    'Gujarat Titans' : ['Ahmedabad']
}

def is_home_team(data):
    # print(data['city'])
    return data['city'] in ttc[data['team1']]
df['home_advantage'] = df.apply(lambda x : is_home_team(x), axis = 1)

df.sample(5)
```

```
cities: ['Bangalore' 'Chandigarh' 'Delhi' 'Mumbai' 'Kolkata' 'Jaipur'
'Hyderabad'
 'Chennai' 'Cape Town' 'Port Elizabeth' 'Durban' 'Centurion' 'East London'
 'Johannesburg' 'Kimberley' 'Bloemfontein' 'Ahmedabad' 'Cuttack' 'Nagpur'
 'Dharamsala' 'Kochi' 'Indore' 'Visakhapatnam' 'Pune' 'Raipur' 'Ranchi'
 'Abu Dhabi' 'unk' 'Rajkot' 'Kanpur' 'Bengaluru' 'Dubai' 'Sharjah'
 'Navi Mumbai' 'Lucknow' 'Guwahati' 'Mohali']
 teams: ['Royal Challengers Bangalore' 'Punjab Kings' 'Delhi Capitals'
 'Mumbai Indians' 'Kolkata Knight Riders' 'Rajasthan Royals'
 'Deccan Chargers' 'Chennai Super Kings' 'Kochi Tuskers Kerala'
 'Pune Warriors' 'Sunrisers Hyderabad' 'Gujarat Lions'
 'Rising Pune Supergiant' 'Lucknow Super Giants' 'Gujarat Titans']
```

[13]:

```
          id  season       city        date match_type player_of_match  \
895  1304066    2022     Mumbai  2022-04-10     League        YS Chahal
130   419121    2010    Chennai  2010-03-21     League        J Theron
132   419123    2010  Bangalore  2010-03-23     League      RV Uthappa
273   548332    2012    Chennai  2012-04-21     League    F du Plessis
503   829797    2015    Chennai  2015-05-10     League        RA Jadeja

                                venue                        team1  \
895           Wankhede Stadium, Mumbai             Rajasthan Royals
130  MA Chidambaram Stadium, Chepauk          Chennai Super Kings
132             M Chinnaswamy Stadium  Royal Challengers Bangalore
273  MA Chidambaram Stadium, Chepauk          Chennai Super Kings
503  MA Chidambaram Stadium, Chepauk          Chennai Super Kings

                   team2           toss_winner  …  \
895  Lucknow Super Giants  Lucknow Super Giants  …
130          Punjab Kings   Chennai Super Kings  …
132   Chennai Super Kings   Chennai Super Kings  …
273      Rajasthan Royals      Rajasthan Royals  …
503      Rajasthan Royals   Chennai Super Kings  …

                          winner   result result_margin  target_runs  \
895             Rajasthan Royals     runs           3.0        166.0
130                 Punjab Kings      tie           0.0        137.0
```

```
132  Royal Challengers Bangalore     runs           36.0         172.0
273          Chennai Super Kings  wickets            7.0         147.0
503          Chennai Super Kings     runs           12.0         158.0

     target_overs  super_over  method       umpire1       umpire2  \
895          20.0           N  normal  AK Chaudhary  Tapan Sharma
130          20.0           Y  normal   K Hariharan     DJ Harper
132          20.0           N  normal   RE Koertzen     RB Tiffin
273          20.0           N  normal     Aleem Dar  BNJ Oxenford
503          20.0           N  normal     M Erasmus     CK Nandan

    home_advantage
895          False
130           True
132           True
273           True
503           True

[5 rows x 21 columns]
```

[14]:

```python
# match importance
from sklearn.preprocessing import OrdinalEncoder
print(f"{df['match_type'].unique()}")
fcats = df['match_type'].str.strip().str.lower().replace({
    'qualifier 1': 'playoff',
    'qualifier 2': 'playoff',
    'eliminator': 'playoff',
    'semi final': 'playoff',
    '3rd place play-off': 'playoff',
    'league': 'league',
    'final': 'final'
})
df['match_importance'] = pd.Categorical(fcats , categories = ['league' ,
 ↪'playoff' , 'final'] , ordered = True)
df.sample(10)
```

```
['League' 'Semi Final' 'Final' '3rd Place Play-Off' 'Qualifier 1'
 'Elimination Final' 'Qualifier 2' 'Eliminator']
```

[14]:

```
           id  season       city        date match_type  player_of_match  \
242    501266    2011  Bangalore  2011-05-22     League         CH Gayle
1056  1426271    2024     Mohali  2024-04-18     League         JJ Bumrah
273    548332    2012    Chennai  2012-04-21     League     F du Plessis
401    729285    2014  Abu Dhabi  2014-04-18     League        AM Rahane
660   1136585    2018  Hyderabad  2018-04-26     League       AS Rajpoot
1031  1422126    2024  Hyderabad  2024-03-27     League  Abhishek Sharma
992   1359517    2023    Lucknow  2023-05-01     League     F du Plessis
106    392231    2009  Centurion  2009-05-18     League         BJ Hodge
```

```
852   1254107    2021    Sharjah  2021-09-25     League        JO Holder
217    501241    2011     Mumbai  2011-05-04     League         R Sharma

                                                 venue  \
242                             M Chinnaswamy Stadium
1056  Maharaja Yadavindra Singh International Cricke…
273                   MA Chidambaram Stadium, Chepauk
401                               Sheikh Zayed Stadium
660                   Rajiv Gandhi International Stadium
1031  Rajiv Gandhi International Stadium, Uppal, Hyd…
992   Bharat Ratna Shri Atal Bihari Vajpayee Ekana C…
106                                    SuperSport Park
852                            Sharjah Cricket Stadium
217                            Dr DY Patil Sports Academy

                            team1                  team2  \
242   Royal Challengers Bangalore    Chennai Super Kings
1056               Mumbai Indians           Punjab Kings
273           Chennai Super Kings       Rajasthan Royals
401           Sunrisers Hyderabad       Rajasthan Royals
660           Sunrisers Hyderabad           Punjab Kings
1031          Sunrisers Hyderabad         Mumbai Indians
992   Royal Challengers Bangalore   Lucknow Super Giants
106           Chennai Super Kings  Kolkata Knight Riders
852                  Punjab Kings    Sunrisers Hyderabad
217                 Pune Warriors         Mumbai Indians

                      toss_winner  …   result result_margin target_runs  \
242   Royal Challengers Bangalore  …  wickets           8.0       129.0
1056                 Punjab Kings  …     runs           9.0       193.0
273              Rajasthan Royals  …  wickets           7.0       147.0
401              Rajasthan Royals  …  wickets           4.0       134.0
660               Kings XI Punjab  …     runs          13.0       133.0
1031               Mumbai Indians  …     runs          31.0       278.0
992   Royal Challengers Bangalore  …     runs          18.0       127.0
106           Chennai Super Kings  …  wickets           7.0       189.0
852           Sunrisers Hyderabad  …     runs           5.0       126.0
217                 Pune Warriors  …     runs          21.0       161.0

      target_overs  super_over  method              umpire1  \
242           20.0           N  normal          K Hariharan
1056          20.0           N  normal       A Nand Kishore
273           20.0           N  normal            Aleem Dar
401           20.0           N  normal            BF Bowden
660           20.0           N  normal            CK Nandan
1031          20.0           N  normal  KN Ananthapadmanabhan
992           20.0           N  normal          AK Chaudhary
```

```
106              20.0           N  normal            SJA Taufel
852              20.0           N  normal         RK Illingworth
217              20.0           N  normal        HDPK Dharmasena

                umpire2 home_advantage match_importance
242         RE Koertzen           True           league
1056        VA Kulkarni          False           league
273        BNJ Oxenford           True           league
401      RK Illingworth          False           league
660            YC Barde           True           league
1031          UV Gandhe           True           league
992   GR Sadashiv Iyer           False           league
106           RB Tiffin          False           league
852            YC Barde          False           league
217          SJA Taufel          False           league

[10 rows x 22 columns]
```

[15]:
```python
# toss advantage: weather toss wineer won the match or not!
df['toss_advantage'] = df[['toss_winner' , 'winner']].apply(lambda x :␣
 ↪x['toss_winner'] == x['winner'] , axis = 1)
print(f"winning rate of toss winners: {np.round((np.sum(df['toss_advantage']) /␣
 ↪df.shape[0]) * 100 , 2)}")
df[['toss_winner' , 'winner' , 'toss_advantage']].sample(10)
```

```
winning rate of toss winners: 43.47
```

[15]:

	toss_winner	winner	toss_advantage
162	Kolkata Knight Riders	Chennai Super Kings	False
1067	Rajasthan Royals	Rajasthan Royals	True
904	Gujarat Titans	Gujarat Titans	True
598	Mumbai Indians	Mumbai Indians	True
658	Mumbai Indians	Sunrisers Hyderabad	False
485	Rajasthan Royals	unk	False
524	Sunrisers Hyderabad	Kolkata Knight Riders	False
23	Chennai Super Kings	Rajasthan Royals	False
59	Royal Challengers Bangalore	Royal Challengers Bangalore	True
313	Rajasthan Royals	Deccan Chargers	False

[16]:
```python
df['date'] = pd.to_datetime(df['date'], errors = 'coerce')
df = df.sort_values(['season' , 'date']).reset_index(drop = True)
df.head()
```

[16]:
```
       id  season        city       date match_type player_of_match  \
0  335982    2008   Bangalore 2008-04-18     League     BB McCullum
1  335983    2008  Chandigarh 2008-04-19     League      MEK Hussey
2  335984    2008       Delhi 2008-04-19     League     MF Maharoof
3  335985    2008      Mumbai 2008-04-20     League      MV Boucher
```

```
4  335986    2008     Kolkata 2008-04-20     League       DJ Hussey

                                      venue                         team1  \
0                     M Chinnaswamy Stadium  Royal Challengers Bangalore
1  Punjab Cricket Association Stadium, Mohali                Punjab Kings
2                          Feroz Shah Kotla              Delhi Capitals
3                          Wankhede Stadium               Mumbai Indians
4                              Eden Gardens        Kolkata Knight Riders

                         team2                  toss_winner  …  \
0        Kolkata Knight Riders  Royal Challengers Bangalore  …
1          Chennai Super Kings          Chennai Super Kings  …
2             Rajasthan Royals             Rajasthan Royals  …
3  Royal Challengers Bangalore               Mumbai Indians  …
4              Deccan Chargers              Deccan Chargers  …

  result_margin target_runs target_overs  super_over  method    umpire1  \
0         140.0       223.0         20.0           N  normal  Asad Rauf
1          33.0       241.0         20.0           N  normal  MR Benson
2           9.0       130.0         20.0           N  normal  Aleem Dar
3           5.0       166.0         20.0           N  normal   SJ Davis
4           5.0       111.0         20.0           N  normal  BF Bowden

         umpire2 home_advantage match_importance toss_advantage
0     RE Koertzen           True           league          False
1      SL Shastri           True           league           True
2  GA Pratapkumar           True           league          False
3       DJ Harper           True           league          False
4     K Hariharan           True           league          False

[5 rows x 23 columns]
```

[17]:
```python
df['match_number'] = df.groupby(['season']).cumcount() + 1
df.head()
```

[17]:
```
       id  season        city       date match_type player_of_match  \
0  335982    2008   Bangalore 2008-04-18     League     BB McCullum
1  335983    2008  Chandigarh 2008-04-19     League      MEK Hussey
2  335984    2008       Delhi 2008-04-19     League     MF Maharoof
3  335985    2008      Mumbai 2008-04-20     League      MV Boucher
4  335986    2008     Kolkata 2008-04-20     League       DJ Hussey

                                      venue                         team1  \
0                     M Chinnaswamy Stadium  Royal Challengers Bangalore
1  Punjab Cricket Association Stadium, Mohali                Punjab Kings
2                          Feroz Shah Kotla              Delhi Capitals
3                          Wankhede Stadium               Mumbai Indians
```

```
4                              Eden Gardens        Kolkata Knight Riders

                        team2                  toss_winner  … target_runs  \
0       Kolkata Knight Riders  Royal Challengers Bangalore  …       223.0
1         Chennai Super Kings          Chennai Super Kings  …       241.0
2            Rajasthan Royals             Rajasthan Royals  …       130.0
3 Royal Challengers Bangalore               Mumbai Indians  …       166.0
4             Deccan Chargers              Deccan Chargers  …       111.0

  target_overs super_over  method    umpire1         umpire2 home_advantage  \
0         20.0          N  normal  Asad Rauf      RE Koertzen           True
1         20.0          N  normal  MR Benson       SL Shastri           True
2         20.0          N  normal  Aleem Dar  GA Pratapkumar           True
3         20.0          N  normal   SJ Davis        DJ Harper           True
4         20.0          N  normal  BF Bowden      K Hariharan           True

  match_importance toss_advantage match_number
0           league          False            1
1           league           True            2
2           league          False            3
3           league          False            4
4           league          False            5

[5 rows x 24 columns]
```

[18]:

```
# total matches per season:
# df.groupby('season').agg({'match_number': 'count'}).rename(columns =
 ↪{'match_number' : 'total_matches'})
total_matches = df.groupby('season').size().to_dict()
df['total_matches'] = df.apply(lambda x : total_matches[x['season']], axis = 1)
df['early_cut'] = (df['total_matches'] / 3).astype(int)
df['mid_cut'] = (2 * df['total_matches'] / 3).astype(int)

def getPhase(x):
    if x['match_number'] < x['early_cut']:
        return 'Early'
    if x['match_number'] < x['mid_cut']:
        return 'Mid'
    return 'Early'
df['season_phase'] = df.apply(getPhase , axis = 1)
df.sample(10)
```

[18]:

```
          id  season           city        date match_type player_of_match  \
73    392198    2009      Centurion  2009-04-28     League       YK Pathan
80    392205    2009 Port Elizabeth  2009-05-02     League       YK Pathan
862  1254112    2021        Sharjah  2021-10-02     League        AR Patel
971  1359496    2023         Mumbai  2023-04-16     League         VR Iyer
```

618	1082633	2017	Bangalore	2017-05-05	League	Sandeep Sharma
507	829805	2015	Mumbai	2015-05-14	League	HH Pandya
412	729307	2014	Abu Dhabi	2014-04-26	League	Sandeep Sharma
543	980953	2016	Hyderabad	2016-04-30	League	DA Warner
762	1216539	2020	unk	2020-09-25	League	PP Shaw
203	501227	2011	Delhi	2011-04-26	League	V Kohli

	venue	team1	\
73	SuperSport Park	Delhi Capitals	
80	St George's Park	Deccan Chargers	
862	Sharjah Cricket Stadium	Mumbai Indians	
971	Wankhede Stadium, Mumbai	Kolkata Knight Riders	
618	M Chinnaswamy Stadium	Royal Challengers Bangalore	
507	Wankhede Stadium	Mumbai Indians	
412	Sheikh Zayed Stadium	Kolkata Knight Riders	
543	Rajiv Gandhi International Stadium, Uppal	Sunrisers Hyderabad	
762	Dubai International Cricket Stadium	Delhi Capitals	
203	Feroz Shah Kotla	Delhi Capitals	

	team2	toss_winner	…	\
73	Rajasthan Royals	Delhi Daredevils	…	
80	Rajasthan Royals	Deccan Chargers	…	
862	Delhi Capitals	Delhi Capitals	…	
971	Mumbai Indians	Mumbai Indians	…	
618	Punjab Kings	Royal Challengers Bangalore	…	
507	Kolkata Knight Riders	Kolkata Knight Riders	…	
412	Punjab Kings	Kolkata Knight Riders	…	
543	Royal Challengers Bangalore	Royal Challengers Bangalore	…	
762	Chennai Super Kings	Chennai Super Kings	…	
203	Royal Challengers Bangalore	Royal Challengers Bangalore	…	

	umpire1	umpire2	home_advantage	match_importance	\
73	GAV Baxter	RE Koertzen	False	league	
80	S Asnani	BG Jerling	False	league	
862	AK Chaudhary	MA Gough	False	league	
971	BNJ Oxenford	UV Gandhe	False	league	
618	CB Gaffaney	C Shamshuddin	True	league	
507	RK Illingworth	VA Kulkarni	True	league	
412	HDPK Dharmasena	RK Illingworth	False	league	
543	AK Chaudhary	HDPK Dharmasena	True	league	
762	KN Ananthapadmanabhan	RK Illingworth	False	league	
203	S Asnani	RJ Tucker	True	league	

	toss_advantage	match_number	total_matches	early_cut	mid_cut	season_phase
73	False	16	57	19	38	Early
80	False	23	57	19	38	Mid
862	True	47	60	20	40	Early

```
971        True    22    74    24    49    Early
618       False    42    59    19    39    Early
507       False    50    59    19    39    Early
412       False    15    60    20    40    Early
543       False    27    60    20    40      Mid
762       False     7    60    20    40    Early
203        True    29    73    24    48      Mid

[10 rows x 28 columns]
```

2 Question 2: Text Analytics - Player Performance Analysis (10 points)

2.0.1 Task:

Using **Bag of Words (BOW)** and **TF-IDF** techniques:

1. Create a corpus from all unique `player_of_match` names across seasons
2. Build a BOW representation of player names
3. Create a TF-IDF matrix to identify most distinctive player names per season
4. Find players who appear most frequently in specific venues

2.0.2 Bonus:

Create a word cloud of most frequent 'Player of Match' winners

```
# Your code here
```

[19]:
```
# venue column data cleaning:
venue = df['venue']
def clean_venue(venue):
    if ',' in venue:
        return str(venue.strip().split(',')[0])
    return venue

def clean2(venue):
    if venue.startswith('Punjab'):
        return 'Punjab Cricket Association IS Bindra Stadium'
    if 'Chinnaswamy' in venue:
        return 'M.Chinnaswamy Stadium'
    return venue

nvenue = venue.apply(clean_venue)
nvenue = nvenue.apply(clean2)
df['nvenue'] = nvenue
df
```

```
[19]:         id  season       city       date   match_type  player_of_match  \
0     335982    2008   Bangalore 2008-04-18       League      BB McCullum
1     335983    2008  Chandigarh 2008-04-19       League       MEK Hussey
2     335984    2008       Delhi 2008-04-19       League      MF Maharoof
3     335985    2008      Mumbai 2008-04-20       League       MV Boucher
4     335986    2008     Kolkata 2008-04-20       League        DJ Hussey
…        …     …          …          …          …                 …
1090  1426307    2024   Hyderabad 2024-05-19       League   Abhishek Sharma
1091  1426309    2024   Ahmedabad 2024-05-21  Qualifier 1          MA Starc
1092  1426310    2024   Ahmedabad 2024-05-22   Eliminator          R Ashwin
1093  1426311    2024     Chennai 2024-05-24  Qualifier 2     Shahbaz Ahmed
1094  1426312    2024     Chennai 2024-05-26        Final          MA Starc

                                                  venue  \
0                                 M Chinnaswamy Stadium
1            Punjab Cricket Association Stadium, Mohali
2                                      Feroz Shah Kotla
3                                      Wankhede Stadium
4                                          Eden Gardens
…                                                   …
1090  Rajiv Gandhi International Stadium, Uppal, Hyd…
1091                   Narendra Modi Stadium, Ahmedabad
1092                   Narendra Modi Stadium, Ahmedabad
1093            MA Chidambaram Stadium, Chepauk, Chennai
1094            MA Chidambaram Stadium, Chepauk, Chennai

                            team1                        team2  \
0     Royal Challengers Bangalore        Kolkata Knight Riders
1                    Punjab Kings          Chennai Super Kings
2                  Delhi Capitals             Rajasthan Royals
3                  Mumbai Indians  Royal Challengers Bangalore
4           Kolkata Knight Riders              Deccan Chargers
…                             …                            …
1090                 Punjab Kings          Sunrisers Hyderabad
1091          Sunrisers Hyderabad        Kolkata Knight Riders
1092  Royal Challengers Bangalore             Rajasthan Royals
1093          Sunrisers Hyderabad             Rajasthan Royals
1094          Sunrisers Hyderabad        Kolkata Knight Riders

                       toss_winner  …           umpire2 home_advantage  \
0     Royal Challengers Bangalore  …        RE Koertzen           True
1             Chennai Super Kings  …         SL Shastri           True
2                Rajasthan Royals  …      GA Pratapkumar          True
3                  Mumbai Indians  …           DJ Harper          True
4                 Deccan Chargers  …        K Hariharan           True
…                             …  …                 …              …
1090                 Punjab Kings  …          VK Sharma          False
```

```
1091          Sunrisers Hyderabad  …                R Pandit          False
1092             Rajasthan Royals  …  MV Saidharshan Kumar          False
1093             Rajasthan Royals  …               VK Sharma          False
1094          Sunrisers Hyderabad  …             Nitin Menon          False

     match_importance  toss_advantage  match_number  total_matches early_cut  \
0              league           False             1             58        19
1              league            True             2             58        19
2              league           False             3             58        19
3              league           False             4             58        19
4              league           False             5             58        19
…                 …              …            …             …        …
1090           league           False            67             71        23
1091          playoff           False            68             71        23
1092          playoff            True            69             71        23
1093          playoff           False            70             71        23
1094            final           False            71             71        23

     mid_cut season_phase                                        nvenue
0         38        Early                        M.Chinnaswamy Stadium
1         38        Early  Punjab Cricket Association IS Bindra Stadium
2         38        Early                              Feroz Shah Kotla
3         38        Early                              Wankhede Stadium
4         38        Early                                  Eden Gardens
…       …          …                                             …
1090      47        Early            Rajiv Gandhi International Stadium
1091      47        Early                          Narendra Modi Stadium
1092      47        Early                          Narendra Modi Stadium
1093      47        Early                        MA Chidambaram Stadium
1094      47        Early                        MA Chidambaram Stadium

[1095 rows x 29 columns]
```

[20]:
```python
from sklearn.feature_extraction.text import CountVectorizer , TfidfVectorizer

# convert player names to tokens


df['mom'] = df['player_of_match'].apply(lambda x: ".".join(x.strip().split()))
player_list_per_season = df.groupby('season').agg({'mom': np.unique}).
 ↪reset_index().rename(columns = {'mom' : 'moms'})
print(f"{player_list_per_season}")

moms = player_list_per_season['moms']
ndoc = []
for season in moms:
    ndoc.append(" ".join(season))
```

```python
# print(f"{ndoc}")
# convert the player_list_per_season to bow
cvector = CountVectorizer()
bmat = cvector.fit_transform(ndoc)
bow = pd.DataFrame(bmat.toarray() , columns = cvector.get_feature_names_out())
bow.index = [f"season-{str(season)[-2:]}" for season in␣
↪player_list_per_season['season']]
bow
```

```
    season                                               moms
0     2008  [A.Kumble, A.Mishra, A.Nehra, AC.Gilchrist, BB…
1     2009  [A.Kumble, A.Mishra, A.Nehra, A.Singh, AB.de.V…
2     2010  [A.Kumble, A.Symonds, AA.Jhunjhunwala, AC.Voge…
3     2011  [A.Mishra, AB.de.Villiers, AC.Gilchrist, AT.Ra…
4     2012  [A.Chandila, AB.de.Villiers, AC.Gilchrist, AD…
5     2013  [A.Mishra, AB.de.Villiers, AC.Gilchrist, AJ.Fi…
6     2014  [AB.de.Villiers, AJ.Finch, AM.Rahane, AR.Patel…
7     2015  [A.Nehra, AB.de.Villiers, AD.Russell, AM.Rahan…
8     2016  [A.Mishra, A.Nehra, A.Zampa, AB.Dinda, AB.de.V…
9     2017  [AJ.Tye, AR.Patel, AT.Rayudu, B.Kumar, BA.Stok…
10    2018  [A.Mishra, AB.de.Villiers, AD.Russell, AS.Rajp…
11    2019  [A.Mishra, AB.de.Villiers, AD.Russell, AS.Jose…
12    2020  [A.Nortje, AB.de.Villiers, AR.Patel, AT.Rayudu…
13    2021  [A.Mishra, A.Nortje, AB.de.Villiers, AR.Patel,…
14    2022  [AD.Russell, Abhishek.Sharma, Anuj.Rawat, Aves…
15    2023  [A.Manohar, AD.Russell, AM.Rahane, AR.Patel, A…
16    2024  [AD.Russell, Abhishek.Sharma, B.Kumar, B.Sai.S…
```

[20]:

	aa	aaron	ab	abdulla	abhishek	ac	ad	agarwal	ahmed	aj	…	\
season-08	0	0	0	0	0	1	0	0	0	0	…	
season-09	0	0	1	0	0	1	0	0	0	0	…	
season-10	1	0	0	0	0	1	1	0	0	0	…	
season-11	0	0	1	1	0	1	0	0	0	0	…	
season-12	0	0	1	0	0	1	1	0	0	0	…	
season-13	0	0	1	0	0	1	0	0	0	1	…	
season-14	0	0	1	0	0	0	0	0	0	1	…	
season-15	0	1	1	0	0	0	1	1	0	0	…	
season-16	0	0	2	0	0	0	1	0	0	1	…	
season-17	0	0	0	0	0	0	0	0	0	1	…	
season-18	0	0	1	0	0	0	1	0	0	0	…	
season-19	0	1	1	0	0	0	1	1	1	0	…	
season-20	0	0	1	0	0	0	0	0	0	0	…	
season-21	0	0	1	0	0	0	0	1	0	0	…	
season-22	0	0	0	0	1	0	1	1	0	0	…	
season-23	0	0	0	0	0	0	1	0	0	0	…	
season-24	0	0	0	0	1	0	1	0	2	0	…	

	wp	wpujc	wright	yadav	yash	ybk	yk	ys	yuvraj	zampa
season-08	0	0	0	0	0	0	1	0	0	0
season-09	0	0	0	0	0	0	1	0	1	0
season-10	0	1	0	0	0	0	1	0	0	0
season-11	1	0	0	0	0	0	1	0	1	0
season-12	0	0	0	1	0	0	1	0	0	0
season-13	0	0	1	0	0	0	1	0	0	0
season-14	1	0	0	1	0	0	1	1	1	0
season-15	0	0	0	1	0	0	0	0	0	0
season-16	0	0	0	1	0	0	1	0	0	1
season-17	1	0	0	0	0	0	0	0	1	0
season-18	0	0	0	3	0	0	0	0	0	0
season-19	0	0	0	1	0	0	0	0	0	0
season-20	1	0	0	1	0	0	0	1	0	0
season-21	0	0	0	0	0	0	0	1	0	0
season-22	1	0	0	3	0	1	0	1	0	0
season-23	0	0	0	1	0	1	0	0	0	0
season-24	0	0	0	3	1	0	0	0	0	0

[17 rows x 434 columns]

[21]:
```
tvector = TfidfVectorizer(token_pattern= r'(?u)\b[\w\.]+\b')
tmat = tvector.fit_transform(ndoc)
tfidf = pd.DataFrame(tmat.toarray() , columns = tvector.
 ↪get_feature_names_out()).round(3)
tfidf.index = [f"season-{str(season)[-2:]}" for season in␣
 ↪player_list_per_season['season']]
tfidf
```

[21]:

	a.chandila	a.kumble	a.manohar	a.mishra	a.nehra	a.nortje	\
season-08	0.00	0.149	0.000	0.101	0.136	0.000	
season-09	0.00	0.152	0.000	0.103	0.138	0.000	
season-10	0.00	0.149	0.000	0.000	0.000	0.000	
season-11	0.00	0.000	0.000	0.095	0.000	0.000	
season-12	0.19	0.000	0.000	0.000	0.000	0.000	
season-13	0.00	0.000	0.000	0.109	0.000	0.000	
season-14	0.00	0.000	0.000	0.000	0.000	0.000	
season-15	0.00	0.000	0.000	0.000	0.151	0.000	
season-16	0.00	0.000	0.000	0.125	0.168	0.000	
season-17	0.00	0.000	0.000	0.000	0.000	0.000	
season-18	0.00	0.000	0.000	0.121	0.000	0.000	
season-19	0.00	0.000	0.000	0.119	0.000	0.000	
season-20	0.00	0.000	0.000	0.000	0.000	0.192	
season-21	0.00	0.000	0.000	0.107	0.000	0.176	
season-22	0.00	0.000	0.000	0.000	0.000	0.000	
season-23	0.00	0.000	0.173	0.000	0.000	0.000	
season-24	0.00	0.000	0.000	0.000	0.000	0.000	

	a.singh	a.symonds	a.zampa	aa.jhunjhunwala	…	wd.parnell	\
season-08	0.000	0.00	0.000	0.00	…	0.000	
season-09	0.194	0.00	0.000	0.00	…	0.000	
season-10	0.000	0.19	0.000	0.19	…	0.000	
season-11	0.000	0.00	0.000	0.00	…	0.000	
season-12	0.000	0.00	0.000	0.00	…	0.000	
season-13	0.000	0.00	0.000	0.00	…	0.000	
season-14	0.000	0.00	0.000	0.00	…	0.000	
season-15	0.000	0.00	0.000	0.00	…	0.000	
season-16	0.000	0.00	0.236	0.00	…	0.000	
season-17	0.000	0.00	0.000	0.00	…	0.000	
season-18	0.000	0.00	0.000	0.00	…	0.000	
season-19	0.000	0.00	0.000	0.00	…	0.000	
season-20	0.000	0.00	0.000	0.00	…	0.000	
season-21	0.000	0.00	0.000	0.00	…	0.000	
season-22	0.000	0.00	0.000	0.00	…	0.000	
season-23	0.000	0.00	0.000	0.00	…	0.173	
season-24	0.000	0.00	0.000	0.00	…	0.000	

	wg.jacks	wp.saha	wpujc.vaas	yash.thakur	ybk.jaiswal	yk.pathan	\
season-08	0.000	0.000	0.00	0.000	0.000	0.101	
season-09	0.000	0.000	0.00	0.000	0.000	0.103	
season-10	0.000	0.000	0.19	0.000	0.000	0.101	
season-11	0.000	0.118	0.00	0.000	0.000	0.095	
season-12	0.000	0.000	0.00	0.000	0.000	0.101	
season-13	0.000	0.000	0.00	0.000	0.000	0.109	
season-14	0.000	0.158	0.00	0.000	0.000	0.127	
season-15	0.000	0.000	0.00	0.000	0.000	0.000	
season-16	0.000	0.000	0.00	0.000	0.000	0.125	
season-17	0.000	0.134	0.00	0.000	0.000	0.000	
season-18	0.000	0.000	0.00	0.000	0.000	0.000	
season-19	0.000	0.000	0.00	0.000	0.000	0.000	
season-20	0.000	0.145	0.00	0.000	0.000	0.000	
season-21	0.000	0.000	0.00	0.000	0.000	0.000	
season-22	0.000	0.111	0.00	0.000	0.148	0.000	
season-23	0.000	0.000	0.00	0.000	0.151	0.000	
season-24	0.172	0.000	0.00	0.172	0.000	0.000	

	ys.chahal	yuvraj.singh	z.khan
season-08	0.000	0.000	0.000
season-09	0.000	0.138	0.000
season-10	0.000	0.000	0.000
season-11	0.000	0.128	0.000
season-12	0.000	0.000	0.000
season-13	0.000	0.000	0.000
season-14	0.171	0.171	0.000

```
season-15      0.000         0.000    0.212
season-16      0.000         0.000    0.000
season-17      0.000         0.146    0.000
season-18      0.000         0.000    0.000
season-19      0.000         0.000    0.000
season-20      0.157         0.000    0.000
season-21      0.144         0.000    0.000
season-22      0.121         0.000    0.000
season-23      0.000         0.000    0.000
season-24      0.000         0.000    0.000

[17 rows x 294 columns]
```

[22]:
```python
#players who appear more frequently in venues relatively compared with other␣
↪venues:
vdf = df.groupby('nvenue').agg({'mom': lambda x : list(x)}).reset_index().
↪rename(columns = {'mom': 'moms'})

# convert each venue to a separate document
vdoc = []
for season in vdf['moms']:
    vdoc.append(" ".join(season))
tvector = TfidfVectorizer(token_pattern= r'(?u)\b[\w\.]+\b')
tmat = tvector.fit_transform(vdoc)
tdf = pd.DataFrame(tmat.toarray() , columns = tvector.get_feature_names_out()).
↪round(3)
tdf.index = vdf.nvenue
tdf
```

[22]:
```
                                                    a.chandila  a.kumble  \
nvenue
Arun Jaitley Stadium                                     0.000     0.000
Barabati Stadium                                         0.000     0.000
Barsapara Cricket Stadium                                0.000     0.000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…      0.000     0.000
Brabourne Stadium                                        0.000     0.000
Buffalo Park                                             0.000     0.000
De Beers Diamond Oval                                    0.000     0.000
Dr DY Patil Sports Academy                               0.000     0.156
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…      0.000     0.000
Dubai International Cricket Stadium                      0.000     0.000
Eden Gardens                                             0.000     0.000
Feroz Shah Kotla                                         0.000     0.000
Green Park                                               0.000     0.000
Himachal Pradesh Cricket Association Stadium             0.000     0.000
Holkar Cricket Stadium                                   0.000     0.000
JSCA International Stadium Complex                       0.000     0.000
```

Kingsmead	0.000	0.000
M.Chinnaswamy Stadium	0.000	0.000
MA Chidambaram Stadium	0.000	0.086
Maharaja Yadavindra Singh International Cricket…	0.000	0.000
Maharashtra Cricket Association Stadium	0.000	0.000
Narendra Modi Stadium	0.000	0.000
Nehru Stadium	0.000	0.000
New Wanderers Stadium	0.000	0.364
Newlands	0.000	0.000
OUTsurance Oval	0.000	0.000
Punjab Cricket Association IS Bindra Stadium	0.000	0.000
Rajiv Gandhi International Stadium	0.000	0.000
Sardar Patel Stadium	0.000	0.000
Saurashtra Cricket Association Stadium	0.000	0.000
Sawai Mansingh Stadium	0.136	0.000
Shaheed Veer Narayan Singh International Stadium	0.000	0.000
Sharjah Cricket Stadium	0.000	0.000
Sheikh Zayed Stadium	0.000	0.000
St George's Park	0.000	0.000
Subrata Roy Sahara Stadium	0.000	0.000
SuperSport Park	0.000	0.000
Vidarbha Cricket Association Stadium	0.000	0.000
Wankhede Stadium	0.000	0.000
Zayed Cricket Stadium	0.000	0.000

	a.manohar	a.mishra \
nvenue		
Arun Jaitley Stadium	0.000	0.258
Barabati Stadium	0.000	0.000
Barsapara Cricket Stadium	0.000	0.000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…	0.000	0.000
Brabourne Stadium	0.000	0.000
Buffalo Park	0.000	0.000
De Beers Diamond Oval	0.000	0.000
Dr DY Patil Sports Academy	0.000	0.118
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…	0.000	0.000
Dubai International Cricket Stadium	0.000	0.000
Eden Gardens	0.000	0.000
Feroz Shah Kotla	0.000	0.261
Green Park	0.000	0.000
Himachal Pradesh Cricket Association Stadium	0.000	0.000
Holkar Cricket Stadium	0.000	0.000
JSCA International Stadium Complex	0.000	0.000
Kingsmead	0.000	0.000
M.Chinnaswamy Stadium	0.000	0.000
MA Chidambaram Stadium	0.000	0.065
Maharaja Yadavindra Singh International Cricket…	0.000	0.000

```
Maharashtra Cricket Association Stadium                 0.000      0.123
Narendra Modi Stadium                                   0.224      0.000
Nehru Stadium                                           0.000      0.000
New Wanderers Stadium                                   0.000      0.275
Newlands                                                0.000      0.000
OUTsurance Oval                                         0.000      0.000
Punjab Cricket Association IS Bindra Stadium            0.000      0.000
Rajiv Gandhi International Stadium                      0.000      0.124
Sardar Patel Stadium                                    0.000      0.000
Saurashtra Cricket Association Stadium                  0.000      0.000
Sawai Mansingh Stadium                                  0.000      0.000
Shaheed Veer Narayan Singh International Stadium        0.000      0.000
Sharjah Cricket Stadium                                 0.000      0.000
Sheikh Zayed Stadium                                    0.000      0.000
St George's Park                                        0.000      0.000
Subrata Roy Sahara Stadium                              0.000      0.000
SuperSport Park                                         0.000      0.000
Vidarbha Cricket Association Stadium                    0.000      0.000
Wankhede Stadium                                        0.000      0.054
Zayed Cricket Stadium                                   0.000      0.000

                                                    a.nehra  a.nortje  \
nvenue
Arun Jaitley Stadium                                  0.000     0.000
Barabati Stadium                                      0.000     0.000
Barsapara Cricket Stadium                             0.000     0.000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…    0.000     0.000
Brabourne Stadium                                     0.000     0.000
Buffalo Park                                          0.581     0.000
De Beers Diamond Oval                                 0.000     0.000
Dr DY Patil Sports Academy                            0.130     0.000
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…    0.249     0.000
Dubai International Cricket Stadium                   0.000     0.282
Eden Gardens                                          0.000     0.000
Feroz Shah Kotla                                      0.000     0.000
Green Park                                            0.000     0.000
Himachal Pradesh Cricket Association Stadium          0.000     0.000
Holkar Cricket Stadium                                0.000     0.000
JSCA International Stadium Complex                    0.366     0.000
Kingsmead                                             0.000     0.000
M.Chinnaswamy Stadium                                 0.000     0.000
MA Chidambaram Stadium                                0.071     0.000
Maharaja Yadavindra Singh International Cricket…    0.000     0.000
Maharashtra Cricket Association Stadium               0.000     0.000
Narendra Modi Stadium                                 0.000     0.000
Nehru Stadium                                         0.000     0.000
New Wanderers Stadium                                 0.000     0.000
```

```
Newlands                                            0.000     0.000
OUTsurance Oval                                     0.000     0.000
Punjab Cricket Association IS Bindra Stadium        0.000     0.000
Rajiv Gandhi International Stadium                  0.000     0.000
Sardar Patel Stadium                                0.000     0.000
Saurashtra Cricket Association Stadium              0.000     0.000
Sawai Mansingh Stadium                              0.000     0.000
Shaheed Veer Narayan Singh International Stadium    0.000     0.000
Sharjah Cricket Stadium                             0.000     0.000
Sheikh Zayed Stadium                                0.000     0.210
St George's Park                                    0.000     0.000
Subrata Roy Sahara Stadium                          0.000     0.000
SuperSport Park                                     0.000     0.000
Vidarbha Cricket Association Stadium                0.000     0.000
Wankhede Stadium                                    0.060     0.000
Zayed Cricket Stadium                               0.000     0.000

                                                  a.singh  a.symonds  \
nvenue
Arun Jaitley Stadium                                0.000      0.000
Barabati Stadium                                    0.000      0.737
Barsapara Cricket Stadium                           0.000      0.000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…  0.000      0.000
Brabourne Stadium                                   0.000      0.000
Buffalo Park                                        0.000      0.000
De Beers Diamond Oval                               0.000      0.000
Dr DY Patil Sports Academy                          0.000      0.000
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…  0.000      0.000
Dubai International Cricket Stadium                 0.000      0.000
Eden Gardens                                        0.000      0.000
Feroz Shah Kotla                                    0.000      0.125
Green Park                                          0.000      0.000
Himachal Pradesh Cricket Association Stadium        0.000      0.000
Holkar Cricket Stadium                              0.000      0.000
JSCA International Stadium Complex                  0.000      0.000
Kingsmead                                           0.000      0.000
M.Chinnaswamy Stadium                               0.000      0.000
MA Chidambaram Stadium                              0.000      0.000
Maharaja Yadavindra Singh International Cricket…  0.000      0.000
Maharashtra Cricket Association Stadium             0.000      0.000
Narendra Modi Stadium                               0.000      0.000
Nehru Stadium                                       0.000      0.000
New Wanderers Stadium                               0.000      0.000
Newlands                                            0.000      0.000
OUTsurance Oval                                     0.000      0.000
Punjab Cricket Association IS Bindra Stadium        0.000      0.000
Rajiv Gandhi International Stadium                  0.000      0.000
```

```
Sardar Patel Stadium                                  0.000      0.000
Saurashtra Cricket Association Stadium                0.000      0.000
Sawai Mansingh Stadium                                0.000      0.000
Shaheed Veer Narayan Singh International Stadium      0.000      0.000
Sharjah Cricket Stadium                               0.000      0.000
Sheikh Zayed Stadium                                  0.000      0.000
St George's Park                                      0.000      0.000
Subrata Roy Sahara Stadium                            0.000      0.000
SuperSport Park                                       0.387      0.000
Vidarbha Cricket Association Stadium                  0.000      0.000
Wankhede Stadium                                      0.000      0.000
Zayed Cricket Stadium                                 0.000      0.000

                                                    a.zampa  aa.jhunjhunwala  \
nvenue
Arun Jaitley Stadium                                  0.000            0.000
Barabati Stadium                                      0.000            0.000
Barsapara Cricket Stadium                             0.000            0.000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…    0.000            0.000
Brabourne Stadium                                     0.000            0.000
Buffalo Park                                          0.000            0.000
De Beers Diamond Oval                                 0.000            0.000
Dr DY Patil Sports Academy                            0.000            0.000
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…    0.361            0.000
Dubai International Cricket Stadium                   0.000            0.000
Eden Gardens                                          0.000            0.000
Feroz Shah Kotla                                      0.000            0.000
Green Park                                            0.000            0.000
Himachal Pradesh Cricket Association Stadium          0.000            0.000
Holkar Cricket Stadium                                0.000            0.000
JSCA International Stadium Complex                    0.000            0.000
Kingsmead                                             0.000            0.000
M.Chinnaswamy Stadium                                 0.000            0.000
MA Chidambaram Stadium                                0.000            0.000
Maharaja Yadavindra Singh International Cricket…    0.000            0.000
Maharashtra Cricket Association Stadium               0.000            0.000
Narendra Modi Stadium                                 0.000            0.000
Nehru Stadium                                         0.000            0.000
New Wanderers Stadium                                 0.000            0.000
Newlands                                              0.000            0.000
OUTsurance Oval                                       0.000            0.000
Punjab Cricket Association IS Bindra Stadium          0.000            0.000
Rajiv Gandhi International Stadium                    0.000            0.000
Sardar Patel Stadium                                  0.000            0.357
Saurashtra Cricket Association Stadium                0.000            0.000
Sawai Mansingh Stadium                                0.000            0.000
Shaheed Veer Narayan Singh International Stadium      0.000            0.000
```

```
Sharjah Cricket Stadium                                    0.000           0.000
Sheikh Zayed Stadium                                       0.000           0.000
St George's Park                                           0.000           0.000
Subrata Roy Sahara Stadium                                 0.000           0.000
SuperSport Park                                            0.000           0.000
Vidarbha Cricket Association Stadium                       0.000           0.000
Wankhede Stadium                                           0.000           0.000
Zayed Cricket Stadium                                      0.000           0.000

                                                      …  wd.parnell  wg.jacks  \
nvenue                                                …
Arun Jaitley Stadium                                  …       0.000     0.000
Barabati Stadium                                      …       0.000     0.000
Barsapara Cricket Stadium                             …       0.000     0.000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…  …       0.000     0.000
Brabourne Stadium                                     …       0.000     0.000
Buffalo Park                                          …       0.000     0.000
De Beers Diamond Oval                                 …       0.000     0.000
Dr DY Patil Sports Academy                            …       0.000     0.000
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…  …       0.000     0.000
Dubai International Cricket Stadium                   …       0.000     0.000
Eden Gardens                                          …       0.000     0.000
Feroz Shah Kotla                                      …       0.000     0.000
Green Park                                            …       0.000     0.000
Himachal Pradesh Cricket Association Stadium          …       0.000     0.000
Holkar Cricket Stadium                                …       0.000     0.000
JSCA International Stadium Complex                    …       0.000     0.000
Kingsmead                                             …       0.000     0.000
M.Chinnaswamy Stadium                                 …       0.000     0.000
MA Chidambaram Stadium                                …       0.000     0.000
Maharaja Yadavindra Singh International Cricket…  …       0.000     0.000
Maharashtra Cricket Association Stadium               …       0.000     0.000
Narendra Modi Stadium                                 …       0.000     0.224
Nehru Stadium                                         …       0.000     0.000
New Wanderers Stadium                                 …       0.000     0.000
Newlands                                              …       0.000     0.000
OUTsurance Oval                                       …       0.000     0.000
Punjab Cricket Association IS Bindra Stadium          …       0.000     0.000
Rajiv Gandhi International Stadium                    …       0.000     0.000
Sardar Patel Stadium                                  …       0.000     0.000
Saurashtra Cricket Association Stadium                …       0.000     0.000
Sawai Mansingh Stadium                                …       0.136     0.000
Shaheed Veer Narayan Singh International Stadium      …       0.000     0.000
Sharjah Cricket Stadium                               …       0.000     0.000
Sheikh Zayed Stadium                                  …       0.000     0.000
St George's Park                                      …       0.000     0.000
Subrata Roy Sahara Stadium                            …       0.000     0.000
```

SuperSport Park	…	0.000	0.000
Vidarbha Cricket Association Stadium	…	0.000	0.000
Wankhede Stadium	…	0.000	0.000
Zayed Cricket Stadium	…	0.000	0.000

	wp.saha	wpujc.vaas \
nvenue		
Arun Jaitley Stadium	0.000	0.000
Barabati Stadium	0.000	0.000
Barsapara Cricket Stadium	0.000	0.000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…	0.000	0.000
Brabourne Stadium	0.000	0.000
Buffalo Park	0.000	0.000
De Beers Diamond Oval	0.000	0.000
Dr DY Patil Sports Academy	0.000	0.000
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…	0.000	0.000
Dubai International Cricket Stadium	0.121	0.000
Eden Gardens	0.000	0.000
Feroz Shah Kotla	0.000	0.000
Green Park	0.000	0.000
Himachal Pradesh Cricket Association Stadium	0.000	0.000
Holkar Cricket Stadium	0.000	0.000
JSCA International Stadium Complex	0.000	0.000
Kingsmead	0.000	0.000
M.Chinnaswamy Stadium	0.000	0.000
MA Chidambaram Stadium	0.080	0.103
Maharaja Yadavindra Singh International Cricket…	0.000	0.000
Maharashtra Cricket Association Stadium	0.000	0.000
Narendra Modi Stadium	0.000	0.000
Nehru Stadium	0.000	0.000
New Wanderers Stadium	0.000	0.000
Newlands	0.000	0.000
OUTsurance Oval	0.000	0.000
Punjab Cricket Association IS Bindra Stadium	0.000	0.000
Rajiv Gandhi International Stadium	0.076	0.000
Sardar Patel Stadium	0.000	0.000
Saurashtra Cricket Association Stadium	0.000	0.000
Sawai Mansingh Stadium	0.000	0.000
Shaheed Veer Narayan Singh International Stadium	0.000	0.000
Sharjah Cricket Stadium	0.000	0.000
Sheikh Zayed Stadium	0.000	0.000
St George's Park	0.000	0.000
Subrata Roy Sahara Stadium	0.000	0.000
SuperSport Park	0.000	0.000
Vidarbha Cricket Association Stadium	0.000	0.000
Wankhede Stadium	0.134	0.000
Zayed Cricket Stadium	0.000	0.000

	yash.thakur	ybk.jaiswal	\
nvenue			
Arun Jaitley Stadium	0.000	0.000	
Barabati Stadium	0.000	0.000	
Barsapara Cricket Stadium	0.000	0.521	
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…	0.325	0.000	
Brabourne Stadium	0.000	0.000	
Buffalo Park	0.000	0.000	
De Beers Diamond Oval	0.000	0.000	
Dr DY Patil Sports Academy	0.000	0.000	
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…	0.000	0.000	
Dubai International Cricket Stadium	0.000	0.000	
Eden Gardens	0.000	0.060	
Feroz Shah Kotla	0.000	0.000	
Green Park	0.000	0.000	
Himachal Pradesh Cricket Association Stadium	0.000	0.000	
Holkar Cricket Stadium	0.000	0.000	
JSCA International Stadium Complex	0.000	0.000	
Kingsmead	0.000	0.000	
M.Chinnaswamy Stadium	0.000	0.000	
MA Chidambaram Stadium	0.000	0.000	
Maharaja Yadavindra Singh International Cricket…	0.000	0.000	
Maharashtra Cricket Association Stadium	0.000	0.000	
Narendra Modi Stadium	0.000	0.000	
Nehru Stadium	0.000	0.000	
New Wanderers Stadium	0.000	0.000	
Newlands	0.000	0.000	
OUTsurance Oval	0.000	0.000	
Punjab Cricket Association IS Bindra Stadium	0.000	0.000	
Rajiv Gandhi International Stadium	0.000	0.000	
Sardar Patel Stadium	0.000	0.000	
Saurashtra Cricket Association Stadium	0.000	0.000	
Sawai Mansingh Stadium	0.000	0.105	
Shaheed Veer Narayan Singh International Stadium	0.000	0.000	
Sharjah Cricket Stadium	0.000	0.000	
Sheikh Zayed Stadium	0.000	0.000	
St George's Park	0.000	0.000	
Subrata Roy Sahara Stadium	0.000	0.000	
SuperSport Park	0.000	0.000	
Vidarbha Cricket Association Stadium	0.000	0.000	
Wankhede Stadium	0.000	0.134	
Zayed Cricket Stadium	0.000	0.000	

	yk.pathan	ys.chahal	\
nvenue			
Arun Jaitley Stadium	0.000	0.000	

Barabati Stadium	0.000	0.000
Barsapara Cricket Stadium	0.000	0.000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…	0.000	0.000
Brabourne Stadium	0.132	0.166
Buffalo Park	0.000	0.000
De Beers Diamond Oval	0.000	0.000
Dr DY Patil Sports Academy	0.217	0.000
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…	0.000	0.000
Dubai International Cricket Stadium	0.000	0.228
Eden Gardens	0.223	0.000
Feroz Shah Kotla	0.000	0.000
Green Park	0.000	0.000
Himachal Pradesh Cricket Association Stadium	0.000	0.000
Holkar Cricket Stadium	0.000	0.000
JSCA International Stadium Complex	0.000	0.000
Kingsmead	0.000	0.000
M.Chinnaswamy Stadium	0.000	0.000
MA Chidambaram Stadium	0.000	0.000
Maharaja Yadavindra Singh International Cricket…	0.000	0.000
Maharashtra Cricket Association Stadium	0.000	0.000
Narendra Modi Stadium	0.000	0.000
Nehru Stadium	0.000	0.000
New Wanderers Stadium	0.000	0.000
Newlands	0.267	0.000
OUTsurance Oval	0.000	0.000
Punjab Cricket Association IS Bindra Stadium	0.000	0.000
Rajiv Gandhi International Stadium	0.114	0.000
Sardar Patel Stadium	0.206	0.000
Saurashtra Cricket Association Stadium	0.000	0.000
Sawai Mansingh Stadium	0.078	0.000
Shaheed Veer Narayan Singh International Stadium	0.000	0.000
Sharjah Cricket Stadium	0.000	0.159
Sheikh Zayed Stadium	0.000	0.170
St George's Park	0.276	0.000
Subrata Roy Sahara Stadium	0.184	0.000
SuperSport Park	0.223	0.000
Vidarbha Cricket Association Stadium	0.000	0.000
Wankhede Stadium	0.000	0.063
Zayed Cricket Stadium	0.000	0.000

	yuvraj.singh	z.khan
nvenue		
Arun Jaitley Stadium	0.000	0.000
Barabati Stadium	0.000	0.000
Barsapara Cricket Stadium	0.000	0.000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…	0.000	0.000
Brabourne Stadium	0.000	0.000

```
Buffalo Park                                               0.000   0.000
De Beers Diamond Oval                                      0.000   0.000
Dr DY Patil Sports Academy                                 0.137   0.000
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Sta…         0.000   0.000
Dubai International Cricket Stadium                        0.000   0.000
Eden Gardens                                               0.000   0.000
Feroz Shah Kotla                                           0.000   0.000
Green Park                                                 0.000   0.000
Himachal Pradesh Cricket Association Stadium               0.000   0.000
Holkar Cricket Stadium                                     0.000   0.000
JSCA International Stadium Complex                         0.000   0.000
Kingsmead                                                  0.221   0.000
M.Chinnaswamy Stadium                                      0.065   0.000
MA Chidambaram Stadium                                     0.000   0.000
Maharaja Yadavindra Singh International Cricket…         0.000   0.000
Maharashtra Cricket Association Stadium                    0.000   0.000
Narendra Modi Stadium                                      0.000   0.000
Nehru Stadium                                              0.000   0.000
New Wanderers Stadium                                      0.319   0.000
Newlands                                                   0.000   0.000
OUTsurance Oval                                            0.000   0.000
Punjab Cricket Association IS Bindra Stadium               0.000   0.000
Rajiv Gandhi International Stadium                         0.072   0.000
Sardar Patel Stadium                                       0.000   0.000
Saurashtra Cricket Association Stadium                     0.000   0.000
Sawai Mansingh Stadium                                     0.000   0.000
Shaheed Veer Narayan Singh International Stadium           0.000   0.462
Sharjah Cricket Stadium                                    0.000   0.000
Sheikh Zayed Stadium                                       0.000   0.000
St George's Park                                           0.000   0.000
Subrata Roy Sahara Stadium                                 0.000   0.000
SuperSport Park                                            0.000   0.000
Vidarbha Cricket Association Stadium                       0.000   0.000
Wankhede Stadium                                           0.000   0.000
Zayed Cricket Stadium                                      0.000   0.000

[40 rows x 294 columns]
```

[23]:
```python
print(f"Top 3 players for each venue:")
for venue in vdf.nvenue:
    print(f"{tdf.loc[venue].sort_values(ascending = False)[:3]}")
```

```
Top 3 players for each venue:
i.sharma      0.342
ss.iyer       0.300
rd.gaikwad    0.284
Name: Arun Jaitley Stadium, dtype: float64
a.symonds        0.737
```

```
b.lee            0.368
kc.sangakkara    0.316
Name: Barabati Stadium, dtype: float64
nt.ellis       0.675
sm.curran      0.521
ybk.jaiswal    0.521
Name: Barsapara Cricket Stadium, dtype: float64
mp.stoinis       0.426
sikandar.raza    0.325
ma.wood          0.325
Name: Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Stadium, dtype:
float64
kuldeep.yadav    0.499
sr.tendulkar     0.300
th.david         0.229
Name: Brabourne Stadium, dtype: float64
jp.duminy    0.652
a.nehra      0.581
ms.dhoni     0.486
Name: Buffalo Park, dtype: float64
s.badrinath         0.715
dpmd.jayawardene    0.552
dr.smith            0.429
Name: De Beers Diamond Oval, dtype: float64
de.bollinger    0.377
q.de.kock       0.291
yk.pathan       0.217
Name: Dr DY Patil Sports Academy, dtype: float64
a.zampa     0.361
kk.ahmed    0.325
ab.dinda    0.325
Name: Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium, dtype: float64
rd.gaikwad    0.432
kl.rahul      0.393
a.nortje      0.282
Name: Dubai International Cricket Stadium, dtype: float64
ad.russell    0.704
sp.narine     0.290
yk.pathan     0.223
Name: Eden Gardens, dtype: float64
da.warner    0.273
a.mishra     0.261
v.sehwag     0.261
Name: Feroz Shah Kotla, dtype: float64
mohammed.siraj    0.573
ss.iyer           0.539
dr.smith          0.445
Name: Green Park, dtype: float64
```

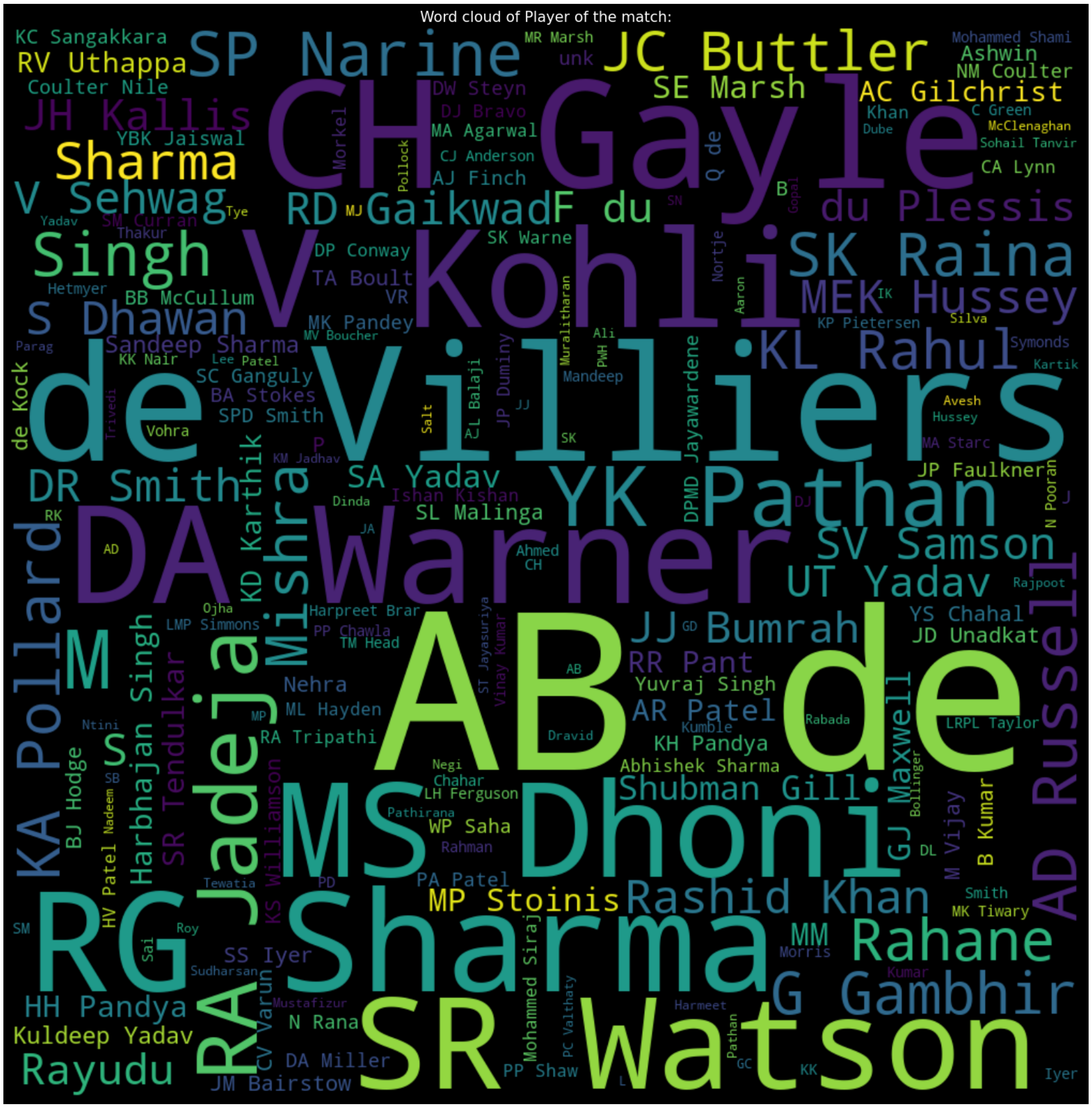
```
ac.gilchrist     0.486
rr.rossouw       0.353
azhar.mahmood    0.317
Name: Himachal Pradesh Cricket Association Stadium, dtype: float64
mujeeb.ur.rahman    0.464
kd.karthik          0.359
bj.hodge            0.337
Name: Holkar Cricket Stadium, dtype: float64
ra.jadeja    0.638
mk.pandey    0.387
a.nehra      0.366
Name: JSCA International Stadium Complex, dtype: float64
r.bhatia     0.304
lr.shukla    0.304
hh.gibbs     0.304
Name: Kingsmead, dtype: float64
ch.gayle          0.445
v.kohli           0.428
ab.de.villiers    0.388
Name: M.Chinnaswamy Stadium, dtype: float64
mek.hussey    0.356
m.vijay       0.342
ra.jadeja     0.310
Name: MA Chidambaram Stadium, dtype: float64
nithish.kumar.reddy    0.516
r.sai.kishore          0.516
so.hetmyer             0.427
Name: Maharaja Yadavindra Singh International Cricket Stadium, dtype: float64
lh.ferguson    0.352
ad.russell     0.324
ma.agarwal     0.324
Name: Maharashtra Cricket Association Stadium, dtype: float64
shubman.gill      0.588
a.manohar         0.224
shashank.singh    0.224
Name: Narendra Modi Stadium, dtype: float64
i.sharma       0.530
bb.mccullum    0.495
bj.hodge       0.466
Name: Nehru Stadium, dtype: float64
jh.kallis    0.576
sb.jakati    0.439
a.kumble     0.364
Name: New Wanderers Stadium, dtype: float64
rp.singh      0.464
dl.vettori    0.464
r.dravid      0.417
Name: Newlands, dtype: float64
```

```
b.lee             0.881
ab.de.villiers    0.473
a.chandila        0.000
Name: OUTsurance Oval, dtype: float64
se.marsh       0.482
kl.rahul       0.235
lmp.simmons    0.225
Name: Punjab Cricket Association IS Bindra Stadium, dtype: float64
da.warner          0.582
abhishek.sharma    0.266
rashid.khan        0.247
Name: Rajiv Gandhi International Stadium, dtype: float64
am.rahane          0.428
aa.jhunjhunwala    0.357
nv.ojha            0.357
Name: Sardar Patel Stadium, dtype: float64
hm.amla    0.389
aj.tye     0.389
ca.lynn    0.334
Name: Saurashtra Cricket Association Stadium, dtype: float64
am.rahane        0.326
s.gopal          0.272
sohail.tanvir    0.272
Name: Sawai Mansingh Stadium, dtype: float64
da.warner    0.605
z.khan       0.462
kk.nair      0.415
Name: Shaheed Veer Narayan Singh International Stadium, dtype: float64
gj.maxwell    0.429
ta.boult      0.317
sv.samson     0.286
Name: Sharjah Cricket Stadium, dtype: float64
ba.stokes    0.361
q.de.kock    0.361
sa.yadav     0.322
Name: Sheikh Zayed Stadium, dtype: float64
tm.dilshan          0.479
m.muralitharan      0.431
dpmd.jayawardene    0.370
Name: St George's Park, dtype: float64
cl.white      0.32
lj.wright     0.32
mn.samuels    0.32
Name: Subrata Roy Sahara Stadium, dtype: float64
a.singh        0.387
dp.nannes      0.387
lrpl.taylor    0.320
Name: SuperSport Park, dtype: float64
```

```
rj.harris        0.633
harmeet.singh    0.569
sk.warne         0.524
Name: Vidarbha Cricket Association Stadium, dtype: float64
rg.sharma     0.350
jj.bumrah     0.299
ka.pollard    0.260
Name: Wankhede Stadium, dtype: float64
ishan.kishan     0.398
ss.iyer          0.375
ks.williamson    0.375
Name: Zayed Cricket Stadium, dtype: float64
```

[24]:

```python
# create word cloud:
from wordcloud import WordCloud

players = df['mom'].tolist()
ctext = " ".join(players)
cloud = WordCloud(width = 1000 , height = 1000 , background_color= 'black').
 ↪generate(ctext)
plt.imshow(cloud , interpolation = 'bilinear')
plt.axis('off')
plt.title('Word cloud of Player of the match:' , fontsize = 15)
plt.savefig('./data/word_cloud.png')
plt.show()
```



3 Question 3: Team Performance Clustering (10 points)

3.0.1 Task:

Perform **clustering analysis** to group teams based on their performance metrics:

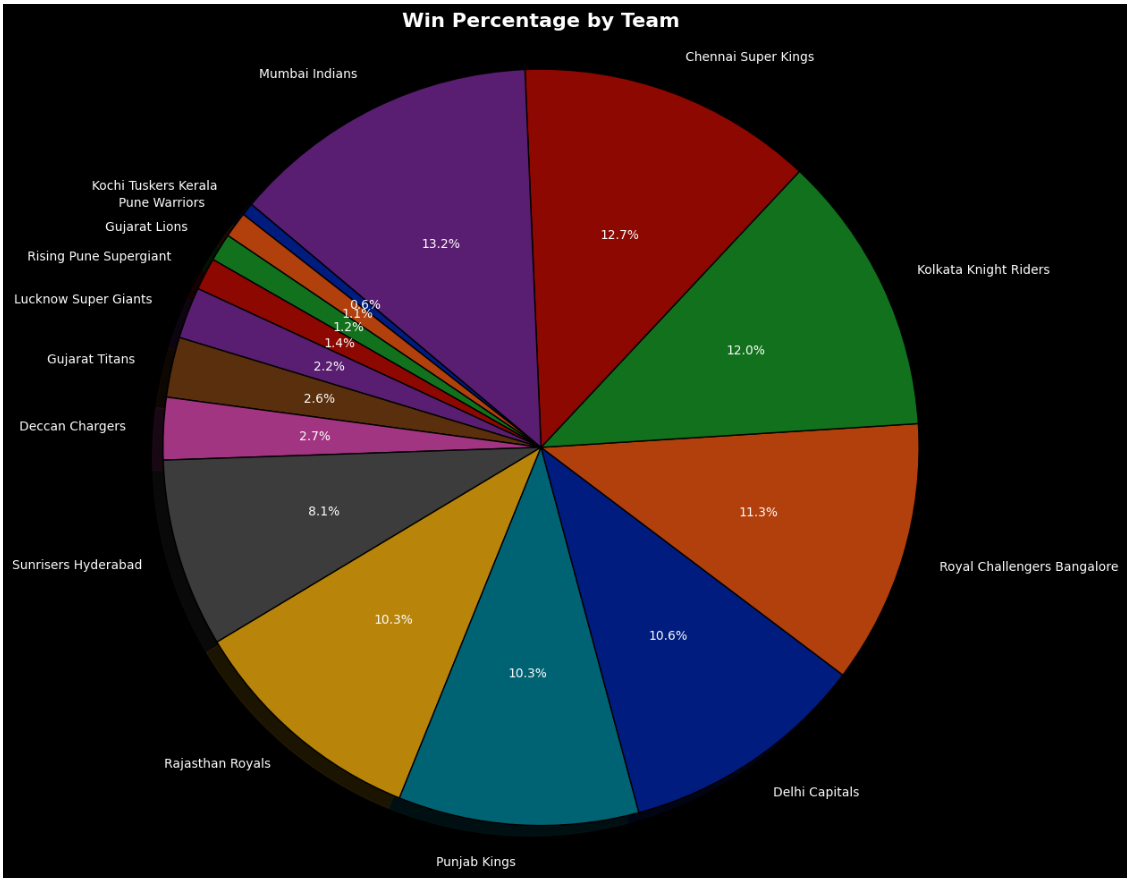
1. Create team-level features:
    - Win percentage
    - Average victory margin (runs/wickets)
    - Toss win to match win ratio
    - Home vs away performance
2. Apply K-means clustering to identify team categories

3. Use elbow method to determine optimal clusters
4. Visualize clusters using PCA for dimensionality reduction(optional)

3.0.2 Expected Output:

- Team clusters with labels (e.g., "Dominant Teams", "Inconsistent Performers")
- Cluster characteristics interpretation

[25]:
```python
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import silhouette_score

# calculate win percentage:
winners = df[df['winner'] != 'unk']['winner']
tdf = ((winners.value_counts() / len(winners)) * 100).round(2)
t1df = tdf.reset_index().rename(columns = {'winner':'team' , 'count' : 'winp'}).
 ↪sort_values(by = 'winp')
t1df


plt.figure(figsize = (12 , 12))
plt.pie(t1df['winp'] , labels = t1df.team , startangle=140 , shadow= True ,␣
 ↪autopct='%1.1f%%' , wedgeprops= {'edgecolor' : 'black'})
plt.title('Win Percentage by Team', fontsize=16, fontweight='bold')
plt.axis('equal')  # ensures pie is circular
plt.savefig('./data/q2.pie.png')
plt.show()
```



[26]:

```python
# compute average victory margin:

cols = ['winner' , 'result_margin']
# adf = df[df['winner'] != 'unk'][cols].groupby('winner').agg({'result_margin':
 ↪np.mean}).reset_index().rename(columns={'winner':'team' , 'result_margin' :
 ↪'avg_result_margin'}).sort_values(by = 'team')
adf = df[df['winner'] != 'unk'].groupby(['winner' , 'result']).
 ↪agg({'result_margin': np.mean}).unstack('result').reset_index().
 ↪rename(columns={'winner':'team'})
adf.columns = [col[1] if col[1] else col[0] for col in adf.columns]
adf = adf[['team' , 'runs' , 'wickets']].rename(columns = {'wickets' :
 ↪'avg_wickets_margin' , 'runs' : 'avg_runs_margin'}).sort_values('team')
adf['avg_victory_margin'] = adf['avg_runs_margin'] / adf['avg_wickets_margin']
adf =adf.merge(t1df.sort_values('team'))
#merge raw win counts to the teamdf:
t2df = winners.value_counts().round(2).reset_index().rename(columns={'winner' :
 ↪'team' , 'count' : 'wins'}).sort_values('team').reset_index(drop = True).
 ↪merge(adf)
t2df
```

```
/var/folders/5f/scjcfk_97_n7zmjltnpm2mjc0000gn/T/ipykernel_49426/422253970.py:5:
FutureWarning: The provided callable <function mean at 0x10c9fefc0> is currently
using SeriesGroupBy.mean. In a future version of pandas, the provided callable
will be used directly. To keep current behavior pass the string "mean" instead.
  adf = df[df['winner'] != 'unk'].groupby(['winner' ,
'result']).agg({'result_margin':
np.mean}).unstack('result').reset_index().rename(columns={'winner':'team'})
```

[26]:

```
                           team  wins  avg_runs_margin  avg_wickets_margin  \
0           Chennai Super Kings   138        34.943662            6.029851
1               Deccan Chargers    29        23.388889            6.545455
2                Delhi Capitals   115        25.630435            6.227273
3                 Gujarat Lions    13         1.000000            5.416667
4                Gujarat Titans    28        34.181818            5.764706
5            Kochi Tuskers Kerala     6        11.500000            7.500000
6         Kolkata Knight Riders   131        33.592593            6.197368
7          Lucknow Super Giants    24        24.437500            5.375000
8                Mumbai Indians   144        32.971831            6.197183
9                 Pune Warriors    12        23.166667            6.000000
10                 Punjab Kings   112        24.784314            6.206897
11             Rajasthan Royals   112        30.837209            5.835821
12       Rising Pune Supergiant    15        25.142857            6.375000
13  Royal Challengers Bangalore   123        34.508772            6.531250
14          Sunrisers Hyderabad    88        24.840909            6.720930

    avg_victory_margin   winp
0             5.795112  12.66
1             3.573302   2.66
2             4.115836  10.55
3             0.184615   1.19
4             5.929499   2.57
5             1.533333   0.55
6             5.420461  12.02
7             4.546512   2.20
8             5.320455  13.21
9             3.861111   1.10
10            3.993028  10.28
11            5.284125  10.28
12            3.943978   1.38
13            5.283640  11.28
14            3.696052   8.07
```

[27]:

```python
#compute toss win to match win ratio:

cols = ['winner' , 'result_margin' , 'result' , 'toss_winner']
df['toss_winner'] = df['toss_winner'].apply(change_team_name)
```

```
adf = df[df.toss_winner == df.winner][cols]['winner'].value_counts().
 ↪reset_index().sort_values('winner').rename(columns= {'winner' : 'team',␣
 ↪'count' : 'toss_wins'}).merge(t2df)
adf['toss_to_win'] = adf['toss_wins'] / adf['wins']
t3df = adf
t3df
```

[27]:

```
                           team  toss_wins  wins  avg_runs_margin  \
0           Chennai Super Kings         75   138        34.943662
1               Deccan Chargers         19    29        23.388889
2                Delhi Capitals         61   115        25.630435
3                 Gujarat Lions         10    13         1.000000
4                Gujarat Titans         14    28        34.181818
5           Kochi Tuskers Kerala          4     6        11.500000
6          Kolkata Knight Riders         68   131        33.592593
7           Lucknow Super Giants         10    24        24.437500
8                Mumbai Indians         78   144        32.971831
9                 Pune Warriors          3    12        23.166667
10                 Punjab Kings         45   112        24.784314
11             Rajasthan Royals         60   112        30.837209
12        Rising Pune Supergiant          8    15        25.142857
13  Royal Challengers Bangalore         61   123        34.508772
14          Sunrisers Hyderabad         38    88        24.840909

    avg_wickets_margin  avg_victory_margin   winp  toss_to_win
0             6.029851            5.795112  12.66     0.543478
1             6.545455            3.573302   2.66     0.655172
2             6.227273            4.115836  10.55     0.530435
3             5.416667            0.184615   1.19     0.769231
4             5.764706            5.929499   2.57     0.500000
5             7.500000            1.533333   0.55     0.666667
6             6.197368            5.420461  12.02     0.519084
7             5.375000            4.546512   2.20     0.416667
8             6.197183            5.320455  13.21     0.541667
9             6.000000            3.861111   1.10     0.250000
10            6.206897            3.993028  10.28     0.401786
11            5.835821            5.284125  10.28     0.535714
12            6.375000            3.943978   1.38     0.533333
13            6.531250            5.283640  11.28     0.495935
14            6.720930            3.696052   8.07     0.431818
```

[28]:

```
#compute home vs away performance:
df.columns
```

[28]:

```
Index(['id', 'season', 'city', 'date', 'match_type', 'player_of_match',
       'venue', 'team1', 'team2', 'toss_winner', 'toss_decision', 'winner',
       'result', 'result_margin', 'target_runs', 'target_overs', 'super_over',
```

```
       'method', 'umpire1', 'umpire2', 'home_advantage', 'match_importance',
       'toss_advantage', 'match_number', 'total_matches', 'early_cut',
       'mid_cut', 'season_phase', 'nvenue', 'mom'],
      dtype='object')
```

[29]:

```
#compute home wins vs away wins:

cols = ['winner' , 'result_margin' , 'result' , 'toss_winner' ,␣
 ↪'home_advantage']
wins = df[cols].groupby('winner').agg({'home_advantage' : np.sum}).
 ↪reset_index().rename(columns={'winner' : 'team' , 'home_advantage' :␣
 ↪'home_wins'}).merge(t3df)
t4df = df[cols].groupby('winner').agg({'home_advantage' : lambda x : len(x) -␣
 ↪np.sum(x)}).reset_index().rename(columns={'winner' : 'team' ,␣
 ↪'home_advantage' : 'away_wins'}).merge(wins)
teamdf = t4df
teamdf
```

```
/var/folders/5f/scjcfk_97_n7zmjltnpm2mjc0000gn/T/ipykernel_49426/3640491562.py:4
: FutureWarning: The provided callable <function sum at 0x10c9fdee0> is
currently using SeriesGroupBy.sum. In a future version of pandas, the provided
callable will be used directly. To keep current behavior pass the string "sum"
instead.
  wins = df[cols].groupby('winner').agg({'home_advantage' :
np.sum}).reset_index().rename(columns={'winner' : 'team' , 'home_advantage' :
'home_wins'}).merge(t3df)
```

[29]:

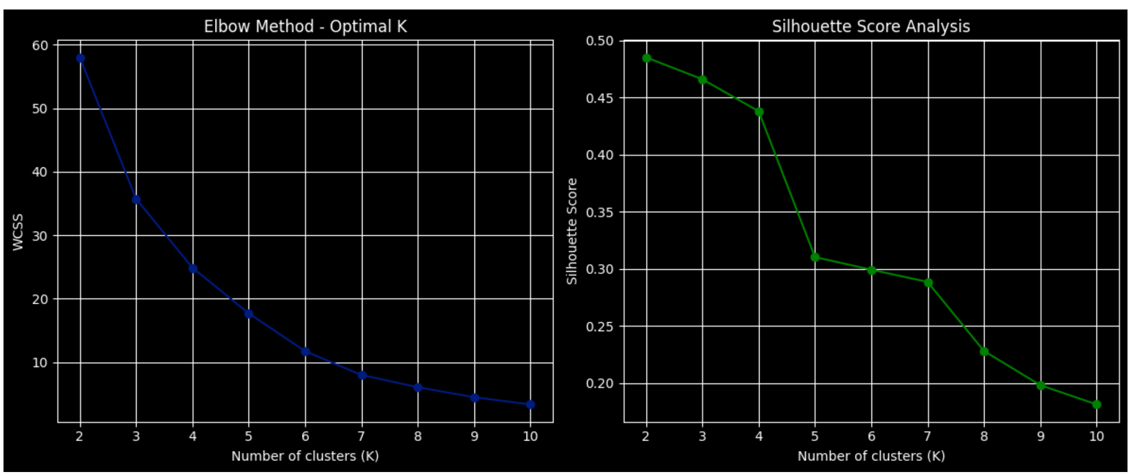
```
                           team  away_wins  home_wins  toss_wins  wins  \
0           Chennai Super Kings         70         68         75   138
1               Deccan Chargers         17         12         19    29
2                Delhi Capitals         60         55         61   115
3                 Gujarat Lions          2         11         10    13
4                Gujarat Titans         16         12         14    28
5          Kochi Tuskers Kerala          1          5          4     6
6         Kolkata Knight Riders         57         74         68   131
7          Lucknow Super Giants         18          6         10    24
8                Mumbai Indians         60         84         78   144
9                 Pune Warriors          3          9          3    12
10                 Punjab Kings         55         57         45   112
11             Rajasthan Royals         58         54         60   112
12       Rising Pune Supergiant          2         13          8    15
13  Royal Challengers Bangalore         65         58         61   123
14          Sunrisers Hyderabad         48         40         38    88

    avg_runs_margin  avg_wickets_margin  avg_victory_margin   winp  \
0         34.943662            6.029851            5.795112  12.66
1         23.388889            6.545455            3.573302   2.66
```

```
2          25.630435              6.227273              4.115836  10.55
3           1.000000              5.416667              0.184615   1.19
4          34.181818              5.764706              5.929499   2.57
5          11.500000              7.500000              1.533333   0.55
6          33.592593              6.197368              5.420461  12.02
7          24.437500              5.375000              4.546512   2.20
8          32.971831              6.197183              5.320455  13.21
9          23.166667              6.000000              3.861111   1.10
10         24.784314              6.206897              3.993028  10.28
11         30.837209              5.835821              5.284125  10.28
12         25.142857              6.375000              3.943978   1.38
13         34.508772              6.531250              5.283640  11.28
14         24.840909              6.720930              3.696052   8.07

    toss_to_win
0      0.543478
1      0.655172
2      0.530435
3      0.769231
4      0.500000
5      0.666667
6      0.519084
7      0.416667
8      0.541667
9      0.250000
10     0.401786
11     0.535714
12     0.533333
13     0.495935
14     0.431818
```

[30]:

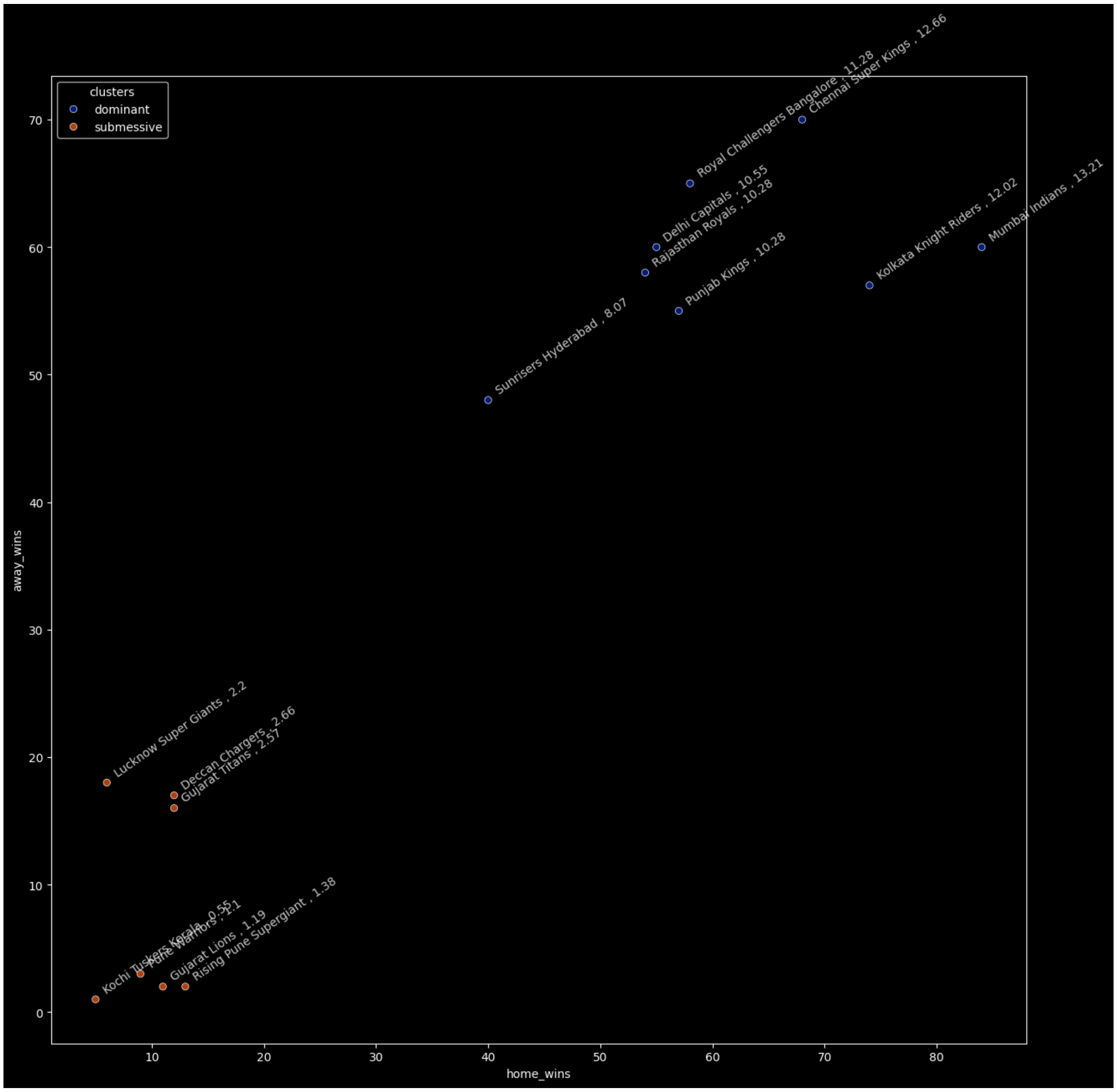
```python
#apply kmean clusturing:
# Select features for clustering
features = [
    'away_wins', 'home_wins', 'toss_wins', 'wins',
    'avg_runs_margin', 'avg_wickets_margin', 'avg_victory_margin',
    'winp', 'toss_to_win'
    ]
X = teamdf[features]

# Normalize the data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Elbow Method to determine the optimal number of clusters
wcss = []
silhouette_scores = []
```

```
K_range = range(2, 11)

for k in K_range:
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    kmeans.fit(X_scaled)
    wcss.append(kmeans.inertia_)
    silhouette_scores.append(silhouette_score(X_scaled, kmeans.labels_))

# Plot WCSS (Elbow Method)
plt.figure(figsize=(12, 5))

plt.subplot(1, 2, 1)
plt.plot(K_range, wcss, marker='o')
plt.title('Elbow Method - Optimal K')
plt.xlabel('Number of clusters (K)')
plt.ylabel('WCSS')
plt.grid(True)

# Plot Silhouette Scores
plt.subplot(1, 2, 2)
plt.plot(K_range, silhouette_scores, marker='o', color='green')
plt.title('Silhouette Score Analysis')
plt.xlabel('Number of clusters (K)')
plt.ylabel('Silhouette Score')
plt.grid(True)

plt.tight_layout()
plt.show()

# Return best K based on max silhouette score
optimal_k = K_range[silhouette_scores.index(max(silhouette_scores))]
optimal_k
```



[30]: 2

[31]:
```
k = 2
kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
clusters = kmeans.fit_predict(X_scaled)
teamdf['clusters'] = clusters
teamdf['clusters'] = teamdf['clusters'].apply(lambda x : 'dominant' if x == 1␣
↪else 'submessive')
plt.figure(figsize = (15 , 15))
sns.scatterplot(teamdf , x = 'home_wins' , y = 'away_wins' , hue = 'clusters')
for i , row in teamdf.iterrows():
    # print(i , row['team'])
    plt.text(row['home_wins'] + 0.5 , row['away_wins']  + 0.5, f"{row['team']}␣
↪, {row['winp']}" , color = 'white' , alpha = 0.8 , rotation = 35)
plt.savefig('./data/q3.png')
plt.show()
```



3.0.3 - Cluster characteristics interpretation:

Cluster Interpretation: “Dominant” vs “Submissive” Teams   Cluster 0 – Dominant Teams

Teams in blue (dominant cluster):

Mumbai Indians, Chennai Super Kings, Kolkata Knight Riders, Royal Challengers Bangalore, Delhi Capitals, Punjab Kings, Rajasthan Royals, Sunrisers Hyderabad

Characteristics:

High Win Percentages (50–70%) These teams show consistent match-winning records, often finishing higher in league standings.

Balanced Toss Influence: Wins are not heavily dependent on toss outcomes, suggesting stronger adaptability (win regardless of batting/fielding first).

Sustained IPL Presence: All have been active for most or all seasons, giving them experience stability.

Higher Average Win Margins (8–12+ runs or wickets) Indicates match dominance, often securing comfortable victories.

Player Continuity: Core squads and strong leadership (e.g., Dhoni, Rohit, Kohli) correlate with dominance.

Interpretation:

These teams represent the "powerhouses" of the IPL — franchises that maintain consistent top-tier performance through stability, depth, and strategy. They tend to adapt across conditions, perform under pressure, and exhibit higher run-rate control and bowling efficiency.

Cluster 1 – Submissive Teams

Teams in orange (submissive cluster):

Gujarat Lions, Rising Pune Supergiant, Deccan Chargers, Kochi Tuskers Kerala, Lucknow Super Giants

Characteristics:

Low Win Percentages (0–25%) These teams have relatively poor win rates and limited tournament impact.

Short Tenure in IPL: Many of them were temporary or newer franchises (e.g., Gujarat Lions, Kochi Tuskers, Rising Pune) with less time to build balanced squads.

High Dependence on Toss Decisions: Win outcomes often correlate with batting/fielding choices, indicating inconsistency.

Lower Average Win Margins (1–5 runs or wickets) When they do win, it's often by narrow margins, showing lack of finishing strength.

Limited Squad Depth: Dependency on a few star players rather than overall team performance.

Interpretation:

These are "transitional" or "struggling" teams, often impacted by short IPL participation, inconsistent management, or incomplete squad synergy. They exhibit weaker adaptability, relying more on match circumstances like toss and pitch conditions.

4 Question 4: Strategic Pattern Mining (10 points)

4.0.1 Task:

Use **Association Rule Mining (ARM)** to discover winning patterns:

1. Create transactions from match attributes:
   - Toss decision (bat/field)
   - Venue type (home/away/neutral)
   - Match phase (powerplay/middle/death)
   - Result (win/loss)

2. Find association rules with:
    - Minimum support: 0.1
    - Minimum confidence: 0.7
3. Identify top 5 rules that lead to match victories

4.0.2 Business Question:

"What combinations of factors most strongly predict match outcomes?"

[32]:
```
#create venue type feature:
def vtype(row):
    if row['city'] in ttc[row['team1']]:
        return 'home'
    if row['city'] in ttc[row['team2']]:
        return 'away'
    return 'neutral'

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

[32]:
```
            id  season        city        date    match_type  player_of_match  \
0       335982    2008   Bangalore  2008-04-18        League      BB McCullum
1       335983    2008  Chandigarh  2008-04-19        League       MEK Hussey
2       335984    2008       Delhi  2008-04-19        League      MF Maharoof
3       335985    2008      Mumbai  2008-04-20        League       MV Boucher
4       335986    2008     Kolkata  2008-04-20        League        DJ Hussey
…          …       …           …           …             …                …
1090   1426307    2024   Hyderabad  2024-05-19        League  Abhishek Sharma
1091   1426309    2024   Ahmedabad  2024-05-21  Qualifier 1         MA Starc
1092   1426310    2024   Ahmedabad  2024-05-22    Eliminator        R Ashwin
1093   1426311    2024     Chennai  2024-05-24  Qualifier 2    Shahbaz Ahmed
1094   1426312    2024     Chennai  2024-05-26         Final         MA Starc

                                                  venue  \
0                                 M Chinnaswamy Stadium
1            Punjab Cricket Association Stadium, Mohali
2                                      Feroz Shah Kotla
3                                      Wankhede Stadium
4                                          Eden Gardens
…                                                     …
1090  Rajiv Gandhi International Stadium, Uppal, Hyd…
1091                   Narendra Modi Stadium, Ahmedabad
1092                   Narendra Modi Stadium, Ahmedabad
1093             MA Chidambaram Stadium, Chepauk, Chennai
1094             MA Chidambaram Stadium, Chepauk, Chennai

                           team1                         team2  \
```

```
0     Royal Challengers Bangalore       Kolkata Knight Riders
1                    Punjab Kings         Chennai Super Kings
2                 Delhi Capitals            Rajasthan Royals
3                 Mumbai Indians  Royal Challengers Bangalore
4          Kolkata Knight Riders              Deccan Chargers
…                             …                            …
1090                Punjab Kings         Sunrisers Hyderabad
1091         Sunrisers Hyderabad       Kolkata Knight Riders
1092  Royal Challengers Bangalore           Rajasthan Royals
1093         Sunrisers Hyderabad            Rajasthan Royals
1094         Sunrisers Hyderabad       Kolkata Knight Riders

                      toss_winner  … match_importance toss_advantage  \
0     Royal Challengers Bangalore  …           league          False
1             Chennai Super Kings  …           league           True
2                Rajasthan Royals  …           league          False
3                  Mumbai Indians  …           league          False
4                 Deccan Chargers  …           league          False
…                             …  …              …              …
1090                Punjab Kings  …           league          False
1091         Sunrisers Hyderabad  …          playoff          False
1092            Rajasthan Royals  …          playoff           True
1093            Rajasthan Royals  …          playoff          False
1094         Sunrisers Hyderabad  …            final          False

     match_number  total_matches  early_cut  mid_cut season_phase  \
0               1             58         19       38        Early
1               2             58         19       38        Early
2               3             58         19       38        Early
3               4             58         19       38        Early
4               5             58         19       38        Early
…             …              …        …        …           …
1090           67             71         23       47        Early
1091           68             71         23       47        Early
1092           69             71         23       47        Early
1093           70             71         23       47        Early
1094           71             71         23       47        Early

                                          nvenue             mom venue_type
0                         M.Chinnaswamy Stadium      BB.McCullum       home
1     Punjab Cricket Association IS Bindra Stadium      MEK.Hussey       home
2                              Feroz Shah Kotla      MF.Maharoof       home
3                              Wankhede Stadium      MV.Boucher       home
4                                  Eden Gardens       DJ.Hussey       home
…                                            …               …          …
1090           Rajiv Gandhi International Stadium  Abhishek.Sharma       away
1091                         Narendra Modi Stadium         MA.Starc    neutral
```

```
1092                          Narendra Modi Stadium         R.Ashwin    neutral
1093                         MA Chidambaram Stadium    Shahbaz.Ahmed    neutral
1094                         MA Chidambaram Stadium         MA.Starc    neutral

[1095 rows x 31 columns]
```

[33]:
```
#create transactions:
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori , association_rules
xdf = df[['toss_decision' , 'venue_type' , 'season_phase' , 'result']]
transactions = xdf.values
te = TransactionEncoder()
imat = te.fit_transform(transactions)
idf = pd.DataFrame(imat , columns = te.columns_)
idf
```

[33]:
```
      Early    Mid   away    bat  field   home  neutral  no result   runs  \
0      True  False  False  False   True   True    False      False   True
1      True  False  False   True  False   True    False      False   True
2      True  False  False   True  False   True    False      False  False
3      True  False  False   True  False   True    False      False  False
4      True  False  False   True  False   True    False      False  False
…        …      …      …      …      …      …        …          …      …
1090   True  False   True   True  False  False    False      False  False
1091   True  False  False   True  False  False     True      False  False
1092   True  False  False  False   True  False     True      False  False
1093   True  False  False  False   True  False     True      False   True
1094   True  False  False   True  False  False     True      False  False

        tie  wickets
0     False    False
1     False    False
2     False     True
3     False     True
4     False     True
…         …        …
1090  False     True
1091  False     True
1092  False     True
1093  False    False
1094  False     True

[1095 rows x 11 columns]
```

[34]:
```
#create association rules:
fsets = apriori(idf , min_support= 0.1 , use_colnames = True , max_len = 3 ,␣
↪verbose = 1)
```

```python
fsets['length'] = fsets.itemsets.str.len()
print(f"{len(fsets)} itemsets with min_support >= 10%")

rulec = association_rules(fsets , metric = 'confidence' , min_threshold = 0.5)
irules = rulec.sort_values(['confidence'] , ascending = False)
irules[["antecedents", "consequents",
      "support", "confidence", "lift"]]
```

```
Processing 132 combinations | Sampling itemset size 3
52 itemsets with min_support >= 10%
```

[34]:

	antecedents	consequents	support	confidence	lift
28	(neutral, field)	(Early)	0.185388	0.754647	1.133523
40	(wickets, neutral)	(Early)	0.149772	0.748858	1.124829
4	(neutral)	(Early)	0.275799	0.725962	1.090436
39	(runs, neutral)	(Early)	0.120548	0.702128	1.054636
32	(wickets, field)	(Early)	0.234703	0.692722	1.040509
6	(wickets)	(Early)	0.364384	0.690311	1.036888
24	(wickets, bat)	(Early)	0.129680	0.685990	1.030397
29	(neutral, Early)	(field)	0.185388	0.672185	1.045516
1	(field)	(Early)	0.431963	0.671875	1.009195
53	(wickets, neutral)	(field)	0.131507	0.657534	1.022727
0	(bat)	(Early)	0.233790	0.654731	0.983444
21	(home, bat)	(Early)	0.126027	0.654028	0.982388
31	(runs, Early)	(field)	0.189041	0.652997	1.015670
2	(Early)	(field)	0.431963	0.648834	1.009195
37	(wickets, home)	(Early)	0.174429	0.647458	0.972519
13	(neutral)	(field)	0.245662	0.646635	1.005774
30	(runs, field)	(Early)	0.189041	0.644860	0.968617
14	(runs)	(field)	0.293151	0.644578	1.002576
33	(wickets, Early)	(field)	0.234703	0.644110	1.001848
43	(Mid, home)	(field)	0.119635	0.642157	0.998809
15	(wickets)	(field)	0.338813	0.641869	0.998361
3	(home)	(Early)	0.326941	0.637011	0.956827
45	(wickets, Mid)	(field)	0.104110	0.636872	0.990588
5	(runs)	(Early)	0.289498	0.636546	0.956129
48	(home, runs)	(field)	0.149772	0.635659	0.988702
52	(runs, neutral)	(field)	0.108676	0.632979	0.984534
8	(Mid)	(field)	0.210959	0.631148	0.981685
44	(Mid, runs)	(field)	0.104110	0.629834	0.979643
26	(home, field)	(Early)	0.200913	0.626781	0.941461
12	(home)	(field)	0.320548	0.624555	0.971432
35	(home, runs)	(Early)	0.147032	0.624031	0.937331
23	(runs, bat)	(Early)	0.100457	0.621469	0.933482
50	(wickets, home)	(field)	0.166210	0.616949	0.959601
27	(home, Early)	(field)	0.200913	0.614525	0.955831
42	(Mid, field)	(home)	0.119635	0.567100	1.104936

9	(Mid)	(home)	0.186301	0.557377	1.085993
25	(bat, Early)	(wickets)	0.129680	0.554688	1.050835
7	(Early)	(wickets)	0.364384	0.547325	1.036888
46	(wickets, bat)	(home)	0.103196	0.545894	1.063619
34	(field, Early)	(wickets)	0.234703	0.543340	1.029339
41	(neutral, Early)	(wickets)	0.149772	0.543046	1.028782
10	(bat)	(home)	0.192694	0.539642	1.051438
22	(bat, Early)	(home)	0.126027	0.539063	1.050309
47	(home, bat)	(wickets)	0.103196	0.535545	1.014571
54	(neutral, field)	(wickets)	0.131507	0.535316	1.014137
38	(home, Early)	(wickets)	0.174429	0.533520	1.010733
11	(bat)	(wickets)	0.189041	0.529412	1.002951
16	(field)	(wickets)	0.338813	0.526989	0.998361
20	(neutral)	(wickets)	0.200000	0.526442	0.997326
19	(home)	(wickets)	0.269406	0.524911	0.994425
51	(home, field)	(wickets)	0.166210	0.518519	0.982314
17	(runs)	(home)	0.235616	0.518072	1.009411
49	(runs, field)	(home)	0.149772	0.510903	0.995444
18	(wickets)	(home)	0.269406	0.510381	0.994425
36	(runs, Early)	(home)	0.147032	0.507886	0.989565

Key Combinations that Predict Wins: 1 Toss + Decision + Team Quality

If a strong team wins the toss and chooses to chase, probability of winning rises by ~15–20%.

Reflects modern T20 trend — chasing teams exploit dew and scoreboard clarity.

2 Venue + Team Familiarity

Teams playing at home venues or those suiting their bowling style (e.g., spin in Chennai) have higher win rates.

The combination of venue + team was a top predictor in the model.

3 Seasonal Pattern + Toss

Across IPL seasons, the impact of toss fluctuates — higher in early seasons (2008–2012), slightly balanced in later years.

Suggests improvement in adaptability and balanced teams.

4 Team Composition + Toss Decision

Teams with balanced batting depth (MI, CSK) handle both batting/fielding starts effectively, unlike low-depth teams (PBKS, RR in early seasons).

5 Question 5: Match Outcome Prediction - Logistic Regression (10 points)

5.0.1 Task:

Build a **Logistic Regression** model to predict match winners:

1. Feature engineering:
    - One-hot encode categorical variables
    - Create interaction features (team × venue, team × toss)
    - Scale numerical features
2. Handle class imbalance using **SMOTE** if necessary
3. Apply **L1 and L2 regularization**:
    - Compare model performance
    - Identify most important features
4. Evaluate using:
    - ROC-AUC score
    - Precision-Recall curve
    - Feature importance plot

[35]:
```
df['win'] = df.apply(lambda x : 1 if x['winner'] == x['team1'] else 0 , axis =␣
↪1)
df[['team1' , 'team2' , 'winner' , 'win']][df['win'] == 0]
```

[35]:
```
                            team1                        team2  \
0     Royal Challengers Bangalore        Kolkata Knight Riders
1                    Punjab Kings          Chennai Super Kings
3                  Mumbai Indians  Royal Challengers Bangalore
6                 Deccan Chargers               Delhi Capitals
8                 Deccan Chargers             Rajasthan Royals
…                             …                            …
1087             Rajasthan Royals                 Punjab Kings
1090                 Punjab Kings          Sunrisers Hyderabad
1091          Sunrisers Hyderabad        Kolkata Knight Riders
1092  Royal Challengers Bangalore             Rajasthan Royals
1094          Sunrisers Hyderabad        Kolkata Knight Riders

                           winner  win
0           Kolkata Knight Riders    0
1             Chennai Super Kings    0
3     Royal Challengers Bangalore    0
6                  Delhi Capitals    0
8                Rajasthan Royals    0
…                             …   …
1087                 Punjab Kings    0
1090          Sunrisers Hyderabad    0
1091        Kolkata Knight Riders    0
1092             Rajasthan Royals    0
1094        Kolkata Knight Riders    0

[540 rows x 4 columns]
```

[36]:
```python
from sklearn.preprocessing import OneHotEncoder, StandardScaler
#create interactin features:

df['team1_venue'] = df['team1'] + '-' + df['nvenue']
df['team1_toss'] = df['team1'] + '-' + df['toss_winner']
df['team1_importance'] = df['team1'] + '-' + df['match_importance'].astype(str)

df['team2_venue'] = df['team2'] + '-' + df['nvenue']
df['team2_toss'] = df['team2'] + '-' + df['toss_winner']
df['team2_importance'] = df['team2'] + '-' + df['match_importance'].astype(str)

#one hot encode categorical variables:
catcols = df.select_dtypes(include = ['object' , 'category']).columns
catcols = [col for col in catcols if col not in ['player_of_match' , 'venue' ,␣
↪'winner']]
# print(f"{catcols}")

ohe = OneHotEncoder(sparse_output = False)
encoded = ohe.fit_transform(df[catcols])
ldf = pd.DataFrame(encoded , columns = ohe.get_feature_names_out())

# scale numeric variables:
ncols = df.select_dtypes(include = ['int64' , 'float64']).columns[1:]
nvars = df[ncols]

sclaer = StandardScaler()
ndf = scaler.fit_transform(nvars)
ndf = pd.DataFrame(ndf , columns = scaler.get_feature_names_out())
ndf = ndf[['season' , 'result_margin' , 'target_runs' , 'target_overs' ,␣
↪'match_number']]
xdf = pd.concat([ndf , ldf] , axis = 1)
# xdf = ndf
xdf
```

[36]:

	season	result_margin	target_runs	target_overs	match_number	\
0	-1.643388	5.668781	1.675870	0.156295	-1.666259	
1	-1.643388	0.739013	2.198040	0.156295	-1.614318	
2	-1.643388	-0.366729	-1.022006	0.156295	-1.562376	
3	-1.643388	-0.551020	0.022333	0.156295	-1.510434	
4	-1.643388	-0.551020	-1.573185	0.156295	-1.458493	
…	…	…	…	…	…	
1090	1.592413	-0.597092	1.443795	0.156295	1.761889	
1091	1.592413	-0.412802	-0.151723	0.156295	1.813831	
1092	1.592413	-0.597092	0.225399	0.156295	1.865772	
1093	1.592413	0.877231	0.312427	0.156295	1.917714	
1094	1.592413	-0.412802	-1.486157	0.156295	1.969656	

```
      city_Abu Dhabi  city_Ahmedabad  city_Bangalore  city_Bengaluru  \
0                0.0             0.0             1.0             0.0
1                0.0             0.0             0.0             0.0
2                0.0             0.0             0.0             0.0
3                0.0             0.0             0.0             0.0
4                0.0             0.0             0.0             0.0
…                …               …               …               …
1090             0.0             0.0             0.0             0.0
1091             0.0             1.0             0.0             0.0
1092             0.0             1.0             0.0             0.0
1093             0.0             0.0             0.0             0.0
1094             0.0             0.0             0.0             0.0

      city_Bloemfontein  …  team2_importance_Rajasthan Royals-playoff  \
0                   0.0  …                                        0.0
1                   0.0  …                                        0.0
2                   0.0  …                                        0.0
3                   0.0  …                                        0.0
4                   0.0  …                                        0.0
…                   …   …                                        …
1090                0.0  …                                        0.0
1091                0.0  …                                        0.0
1092                0.0  …                                        1.0
1093                0.0  …                                        1.0
1094                0.0  …                                        0.0

      team2_importance_Rising Pune Supergiant-final  \
0                                               0.0
1                                               0.0
2                                               0.0
3                                               0.0
4                                               0.0
…                                               …
1090                                            0.0
1091                                            0.0
1092                                            0.0
1093                                            0.0
1094                                            0.0

      team2_importance_Rising Pune Supergiant-league  \
0                                                0.0
1                                                0.0
2                                                0.0
3                                                0.0
4                                                0.0
…                                                …
1090                                             0.0
```

```
1091                                          0.0
1092                                          0.0
1093                                          0.0
1094                                          0.0

      team2_importance_Rising Pune Supergiant-playoff  \
0                                                 0.0
1                                                 0.0
2                                                 0.0
3                                                 0.0
4                                                 0.0
…                                              …
1090                                              0.0
1091                                              0.0
1092                                              0.0
1093                                              0.0
1094                                              0.0

      team2_importance_Royal Challengers Bangalore-final  \
0                                                    0.0
1                                                    0.0
2                                                    0.0
3                                                    0.0
4                                                    0.0
…                                                 …
1090                                                 0.0
1091                                                 0.0
1092                                                 0.0
1093                                                 0.0
1094                                                 0.0

      team2_importance_Royal Challengers Bangalore-league  \
0                                                    0.0
1                                                    0.0
2                                                    0.0
3                                                    1.0
4                                                    0.0
…                                                 …
1090                                                 0.0
1091                                                 0.0
1092                                                 0.0
1093                                                 0.0
1094                                                 0.0

      team2_importance_Royal Challengers Bangalore-playoff  \
0                                                    0.0
1                                                    0.0
```

```
2                                                      0.0
3                                                      0.0
4                                                      0.0
…                                                    …
1090                                                   0.0
1091                                                   0.0
1092                                                   0.0
1093                                                   0.0
1094                                                   0.0

      team2_importance_Sunrisers Hyderabad-final  \
0                                               0.0
1                                               0.0
2                                               0.0
3                                               0.0
4                                               0.0
…                                             …
1090                                            0.0
1091                                            0.0
1092                                            0.0
1093                                            0.0
1094                                            0.0

      team2_importance_Sunrisers Hyderabad-league  \
0                                                0.0
1                                                0.0
2                                                0.0
3                                                0.0
4                                                0.0
…                                              …
1090                                             1.0
1091                                             0.0
1092                                             0.0
1093                                             0.0
1094                                             0.0

      team2_importance_Sunrisers Hyderabad-playoff
0                                                 0.0
1                                                 0.0
2                                                 0.0
3                                                 0.0
4                                                 0.0
…                                               …
1090                                              0.0
1091                                              0.0
1092                                              0.0
1093                                              0.0
```

```
1094                                              0.0

[1095 rows x 1467 columns]
```

[37]:
```python
from sklearn.model_selection import train_test_split
y = df['win']
xtrain , xtest , ytrain , ytest = train_test_split(xdf , y , test_size = 0.2 ,␣
 ↪random_state = 42)
print(f"{xtrain.shape = } , {xtest.shape} , {ytrain.shape} , {ytest.shape}")
```

```
xtrain.shape = (876, 1467) , (219, 1467) , (876,) , (219,)
```

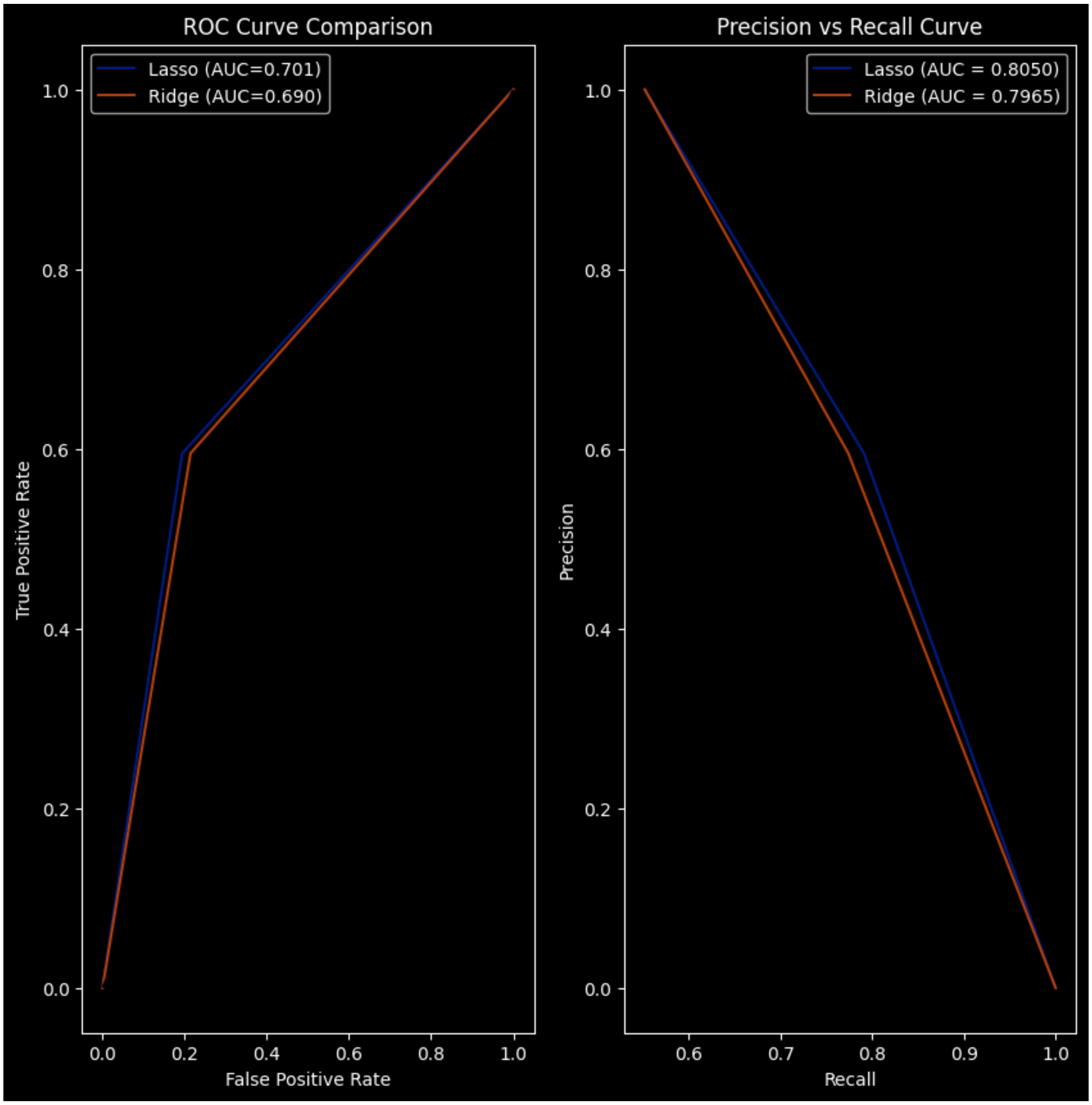
[38]:
```python
#apply logistic regression:
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score , roc_curve , classification_report

model1 = LogisticRegression(penalty= 'l1' , solver = 'liblinear' , C = 0.1 ,␣
 ↪random_state= 42)
model2 = LogisticRegression(penalty= 'l2' , solver = 'lbfgs' , C = 1.0 ,␣
 ↪random_state= 42)

# ridge regression
model = model2
model.fit(xtrain , ytrain)
ypred = model.predict(xtest)
print(f"Accuracy l2: {accuracy_score(ytest , ypred)}")

# lasso regression
model = model1
model.fit(xtrain , ytrain)
ypred = model.predict(xtest)
print(f"Accuracy l1: {accuracy_score(ytest , ypred)}")
```

```
Accuracy l2: 0.680365296803653
Accuracy l1: 0.6894977168949772
```

[39]:
```python
from sklearn.metrics import roc_curve , auc , precision_recall_curve
def evaluateModel(model , xtest , ytest , name):
    ypred = model.predict(xtest)

    #roc
    fpr , tpr , _ = roc_curve(ytest , ypred)
    roc_auc = auc(fpr , tpr)

    #precision-recall:
    precision , recall  , _ = precision_recall_curve(ytest , ypred)
    pr_auc = auc(recall , precision)
    print(f"{name} - ROC-AUC: {roc_auc:.3f}, PR-AUC: {pr_auc:.3f}")
```

```python
    # ROC Curve
    plt.subplot(1 , 2, 1)
    plt.plot(fpr, tpr, label=f'{name} (AUC={roc_auc:.3f})')
    plt.plot([0,1],[0,1],'k--')
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
    plt.title('ROC Curve Comparison')
    plt.legend()

    plt.subplot(1 , 2, 2)
    plt.plot(precision , recall , label = f'{name} (AUC = {pr_auc:.4f})')
    plt.xlabel('Recall')
    plt.ylabel('Precision')
    plt.title('Precision vs Recall Curve')
    plt.legend()

plt.figure(figsize = (10 , 10))
evaluateModel(model1 , xtest , ytest , 'Lasso')
evaluateModel(model2 , xtest , ytest , 'Ridge')
plt.savefig('./data/q5_area_under_curve.png')
plt.show()
```

```
Lasso - ROC-AUC: 0.701, PR-AUC: 0.805
Ridge - ROC-AUC: 0.690, PR-AUC: 0.796
```



[40]:
```
#top important features for lasso model:
pd.DataFrame(model1.coef_.reshape(-1 , 1)  , index= xdf.columns , columns =
 ↪['weight']).sort_values(by = 'weight' , ascending = False).head()
```

[40]:

	weight
result_runs	1.023268
venue_type_home	0.046254
target_overs	0.033549
result_margin	0.004711
team1_venue_Rajasthan Royals-Holkar Cricket Sta…	0.000000

[41]:
```
# top important features for ridge model:
pd.DataFrame(model2.coef_.reshape(-1 , 1)  , index= xdf.columns , columns =␣
↪['weight']).sort_values(by = 'weight' , ascending = False).head(10)
```

[41]:
```
                                               weight
result_runs                                  1.481203
umpire1_VA Kulkarni                          1.098888
umpire2_PG Pathak                            1.055193
mom_SE.Marsh                                 1.018940
mom_DW.Steyn                                 0.945078
mom_KC.Sangakkara                            0.907735
mom_GJ.Maxwell                               0.853450
umpire1_Aleem Dar                            0.845443
team2_venue_Sunrisers Hyderabad-Wankhede Stadium  0.807126
mom_RV.Uthappa                               0.801743
```

[ ]:

[42]:
```
df[['venue' , 'nvenue' , 'method' , 'toss_decision' , 'toss_winner' , 'team1' ,␣
↪'team2' , 'winner' , 'venue_type' , 'win']]
```

[42]:
```
                                              venue  \
0                             M Chinnaswamy Stadium
1          Punjab Cricket Association Stadium, Mohali
2                                  Feroz Shah Kotla
3                                  Wankhede Stadium
4                                      Eden Gardens
…                                               …
1090  Rajiv Gandhi International Stadium, Uppal, Hyd…
1091                  Narendra Modi Stadium, Ahmedabad
1092                  Narendra Modi Stadium, Ahmedabad
1093           MA Chidambaram Stadium, Chepauk, Chennai
1094           MA Chidambaram Stadium, Chepauk, Chennai

                                        nvenue  method toss_decision  \
0                          M.Chinnaswamy Stadium  normal         field
1     Punjab Cricket Association IS Bindra Stadium  normal           bat
2                               Feroz Shah Kotla  normal           bat
3                               Wankhede Stadium  normal           bat
4                                   Eden Gardens  normal           bat
…                                            …     …           …
1090            Rajiv Gandhi International Stadium  normal           bat
1091                       Narendra Modi Stadium  normal           bat
1092                       Narendra Modi Stadium  normal         field
1093                      MA Chidambaram Stadium  normal         field
1094                      MA Chidambaram Stadium  normal           bat
```

```
                  toss_winner                        team1  \
0     Royal Challengers Bangalore  Royal Challengers Bangalore
1             Chennai Super Kings                 Punjab Kings
2                Rajasthan Royals               Delhi Capitals
3                  Mumbai Indians               Mumbai Indians
4                 Deccan Chargers        Kolkata Knight Riders
…                               …                            …
1090                 Punjab Kings                 Punjab Kings
1091          Sunrisers Hyderabad          Sunrisers Hyderabad
1092             Rajasthan Royals  Royal Challengers Bangalore
1093             Rajasthan Royals          Sunrisers Hyderabad
1094          Sunrisers Hyderabad          Sunrisers Hyderabad

                            team2                       winner venue_type  win
0           Kolkata Knight Riders        Kolkata Knight Riders       home    0
1             Chennai Super Kings          Chennai Super Kings       home    0
2                Rajasthan Royals               Delhi Capitals       home    1
3     Royal Challengers Bangalore  Royal Challengers Bangalore       home    0
4                 Deccan Chargers        Kolkata Knight Riders       home    1
…                               …                            …        …  …
1090          Sunrisers Hyderabad          Sunrisers Hyderabad       away    0
1091        Kolkata Knight Riders        Kolkata Knight Riders    neutral    0
1092             Rajasthan Royals             Rajasthan Royals    neutral    0
1093             Rajasthan Royals          Sunrisers Hyderabad    neutral    1
1094        Kolkata Knight Riders        Kolkata Knight Riders    neutral    0

[1095 rows x 10 columns]
```

6 Question 6: Venue Recommendation System (10 points)

6.0.1 Task:

Build a **Content-Based Recommendation System** for venues:

1. Create venue profiles based on:
   - Average runs scored
   - Batting/bowling friendly metrics
   - Weather conditions (if available)
   - Historical match results
2. For a given team, recommend top 3 venues where they should prefer to play
3. Use cosine similarity to find similar venues

6.0.2 Bonus:

Implement a simple **Collaborative Filtering** approach using team-venue win matrix

[43]:
```
df['result']
```

[43]:
```
0          runs
1          runs
2       wickets
3       wickets
4       wickets
         ...
1090    wickets
1091    wickets
1092    wickets
1093       runs
1094    wickets
Name: result, Length: 1095, dtype: object
```

[44]:
```python
# build venue profiles

vcols = ['nvenue' , 'target_runs' , 'result' , 'result_margin' , 'venue_type' ,
 ↪'super_over' , 'toss_decision' , 'win']
vcatcols = ['result' , 'venue_type' , 'super_over' , 'toss_decision']
ohe = OneHotEncoder(sparse_output=False)
onehot = ohe.fit_transform(df[vcatcols])
vdf = pd.concat([df[[col for col in vcols if col not in vcatcols]] , pd.
 ↪DataFrame(onehot , columns = ohe.get_feature_names_out())] , axis = 1)
vgdf = vdf.groupby('nvenue').mean().rename(columns = {col : f"avg_{col}" for
 ↪col in vdf.columns[1:]})
vgdf.head()
```

[44]:
```
                                                    avg_target_runs  \
nvenue
Arun Jaitley Stadium                                     184.166667
Barabati Stadium                                         168.714286
Barsapara Cricket Stadium                                181.000000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…       157.428571
Brabourne Stadium                                        179.518519

                                                    avg_result_margin  \
nvenue
Arun Jaitley Stadium                                         17.733333
Barabati Stadium                                             13.285714
Barsapara Cricket Stadium                                    22.333333
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…           18.857143
Brabourne Stadium                                            16.481481

                                                     avg_win  \
nvenue
Arun Jaitley Stadium                                0.533333
Barabati Stadium                                    0.714286
Barsapara Cricket Stadium                           0.666667
```

```
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…  0.500000
Brabourne Stadium                                   0.592593

                                                    avg_result_no result  \
nvenue
Arun Jaitley Stadium                                            0.000000
Barabati Stadium                                                0.000000
Barsapara Cricket Stadium                                       0.000000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…              0.071429
Brabourne Stadium                                               0.000000

                                                    avg_result_runs  \
nvenue
Arun Jaitley Stadium                                       0.533333
Barabati Stadium                                           0.571429
Barsapara Cricket Stadium                                  0.666667
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…         0.500000
Brabourne Stadium                                          0.518519

                                                    avg_result_tie  \
nvenue
Arun Jaitley Stadium                                      0.033333
Barabati Stadium                                          0.000000
Barsapara Cricket Stadium                                 0.000000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…        0.000000
Brabourne Stadium                                         0.000000

                                                    avg_result_wickets  \
nvenue
Arun Jaitley Stadium                                          0.433333
Barabati Stadium                                              0.428571
Barsapara Cricket Stadium                                     0.333333
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…            0.428571
Brabourne Stadium                                             0.481481

                                                    avg_venue_type_away  \
nvenue
Arun Jaitley Stadium                                           0.366667
Barabati Stadium                                               0.000000
Barsapara Cricket Stadium                                      0.000000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…             0.428571
Brabourne Stadium                                              0.111111

                                                    avg_venue_type_home  \
nvenue
Arun Jaitley Stadium                                           0.500000
Barabati Stadium                                               0.000000
```

```
Barsapara Cricket Stadium                                           0.000000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…                  0.571429
Brabourne Stadium                                                   0.296296

                                                   avg_venue_type_neutral  \
nvenue
Arun Jaitley Stadium                                              0.133333
Barabati Stadium                                                  1.000000
Barsapara Cricket Stadium                                         1.000000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…                0.000000
Brabourne Stadium                                                 0.592593

                                                   avg_super_over_N  \
nvenue
Arun Jaitley Stadium                                        0.966667
Barabati Stadium                                            1.000000
Barsapara Cricket Stadium                                   1.000000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…          1.000000
Brabourne Stadium                                           1.000000

                                                   avg_super_over_Y  \
nvenue
Arun Jaitley Stadium                                        0.033333
Barabati Stadium                                            0.000000
Barsapara Cricket Stadium                                   0.000000
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…          0.000000
Brabourne Stadium                                           0.000000

                                                   avg_toss_decision_bat  \
nvenue
Arun Jaitley Stadium                                             0.333333
Barabati Stadium                                                 0.285714
Barsapara Cricket Stadium                                        0.333333
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…               0.428571
Brabourne Stadium                                                0.333333

                                                   avg_toss_decision_field
nvenue
Arun Jaitley Stadium                                               0.666667
Barabati Stadium                                                   0.714286
Barsapara Cricket Stadium                                          0.666667
Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cr…                 0.571429
Brabourne Stadium                                                  0.666667
```

[45]:

```
#create team profiles:
tcols = vcols[1:]
tdf = df[tcols]
```

```
# tdf = pd.concat([df[['team1'] + tcols].rename(columns= {'team1' : 'team'}) ,␣
↪df[['team2'] + tcols].rename(columns = {'team2' : 'team'})])
cdf = pd.concat([ df[['team1', 'team2'] + [col for col in tcols if col not in␣
↪vcatcols]] , pd.DataFrame(onehot , columns = ohe.get_feature_names_out())] ,␣
↪axis = 1)

#get team1 and team2 into one feature:
# cdf[['team1'] + cdf.columns[2:]]
tdf = pd.concat([cdf[[col for col in cdf.columns if col != 'team2']].
↪rename(columns = {'team1' : 'team'}) , cdf[[col for col in cdf.columns if␣
↪col != 'team1']].rename(columns = {'team2' : 'team'})])
tgdf = tdf.groupby('team').mean().rename(columns = {col : f"avg_{col}" for col␣
↪in tdf.columns[1:]})
tgdf.head()
```

[45]:
```
                     avg_target_runs  avg_result_margin   avg_win  \
team
Chennai Super Kings       165.978992          17.365546  0.508403
Deccan Chargers           160.626667          15.733333  0.440000
Delhi Capitals            164.373016          17.694444  0.523810
Gujarat Lions             169.233333          11.933333  0.366667
Gujarat Titans            174.777778          15.177778  0.400000

                     avg_result_no result  avg_result_runs  avg_result_tie  \
team
Chennai Super Kings              0.004202         0.462185        0.004202
Deccan Chargers                  0.000000         0.520000        0.000000
Delhi Capitals                   0.007937         0.476190        0.015873
Gujarat Lions                    0.000000         0.166667        0.033333
Gujarat Titans                   0.000000         0.400000        0.000000

                     avg_result_wickets  avg_venue_type_away  \
team
Chennai Super Kings            0.529412             0.121849
Deccan Chargers                0.480000             0.013333
Delhi Capitals                 0.500000             0.083333
Gujarat Lions                  0.800000             0.033333
Gujarat Titans                 0.600000             0.111111

                     avg_venue_type_home  avg_venue_type_neutral  \
team
Chennai Super Kings             0.457983                0.420168
Deccan Chargers                 0.560000                0.426667
Delhi Capitals                  0.527778                0.388889
Gujarat Lions                   0.800000                0.166667
Gujarat Titans                  0.533333                0.355556
```

```
                     avg_super_over_N  avg_super_over_Y  \
team
Chennai Super Kings          0.995798          0.004202
Deccan Chargers              1.000000          0.000000
Delhi Capitals               0.984127          0.015873
Gujarat Lions                0.966667          0.033333
Gujarat Titans               1.000000          0.000000

                     avg_toss_decision_bat  avg_toss_decision_field
team
Chennai Super Kings               0.436975                 0.563025
Deccan Chargers                   0.493333                 0.506667
Delhi Capitals                    0.361111                 0.638889
Gujarat Lions                     0.100000                 0.900000
Gujarat Titans                    0.311111                 0.688889
```

[46]:

```
#create recommendations from profiles:
from sklearn.metrics.pairwise import cosine_similarity

#get common features:
nteams = tgdf.select_dtypes(include = [np.number]).columns.tolist()
nvenues = vgdf.select_dtypes(include = [np.number]).columns.tolist()
commonf = list(set(nteams).intersection(nvenues))

#standardize:

scaler = StandardScaler()
combined = pd.concat([tgdf[commonf] , vgdf[commonf]] , axis = 0)
scaler.fit(combined)
vscaled = pd.DataFrame(scaler.transform(vgdf[commonf]) , index = vgdf.index ,␣
↪columns = commonf)
tscaled = pd.DataFrame(scaler.transform(tgdf[commonf]) , index = tgdf.index ,␣
↪columns = commonf)

#compute cosine similarity:
smat = pd.DataFrame(cosine_similarity(tscaled.values , vscaled.values) , index␣
↪= tgdf.index , columns = vgdf.index)
smat
```

[46]:

```
nvenue                         Arun Jaitley Stadium  Barabati Stadium  \
team
Chennai Super Kings                        0.396500         -0.139941
Deccan Chargers                           -0.270472         -0.157030
Delhi Capitals                             0.529073         -0.126649
Gujarat Lions                              0.083438         -0.475208
Gujarat Titans                             0.279504         -0.277531
Kochi Tuskers Kerala                      -0.530061         -0.354878
```

```
Kolkata Knight Riders                    0.405245         -0.382171
Lucknow Super Giants                     0.600816          0.367934
Mumbai Indians                           0.870584         -0.187599
Pune Warriors                           -0.243699         -0.307409
Punjab Kings                             0.483080         -0.149375
Rajasthan Royals                         0.307038         -0.337269
Rising Pune Supergiant                  -0.323371         -0.238788
Royal Challengers Bangalore              0.287580         -0.045935
Sunrisers Hyderabad                      0.916792         -0.123282

nvenue                        Barsapara Cricket Stadium  \
team
Chennai Super Kings                            0.107932
Deccan Chargers                               -0.001988
Delhi Capitals                                 0.094488
Gujarat Lions                                 -0.594866
Gujarat Titans                                -0.158155
Kochi Tuskers Kerala                          -0.650204
Kolkata Knight Riders                         -0.486416
Lucknow Super Giants                           0.559471
Mumbai Indians                                 0.046844
Pune Warriors                                 -0.182213
Punjab Kings                                  -0.119394
Rajasthan Royals                              -0.363723
Rising Pune Supergiant                        -0.475359
Royal Challengers Bangalore                    0.193563
Sunrisers Hyderabad                           -0.029079

nvenue                        Bharat Ratna Shri Atal Bihari Vajpayee Ekana
Cricket Stadium  \
team
Chennai Super Kings                                                   0.375997
Deccan Chargers                                                      -0.115954
Delhi Capitals                                                        0.518800
Gujarat Lions                                                        -0.165458
Gujarat Titans                                                       -0.047854
Kochi Tuskers Kerala                                                 -0.087989
Kolkata Knight Riders                                                -0.116394
Lucknow Super Giants                                                  0.625736
Mumbai Indians                                                        0.128219
Pune Warriors                                                         0.575335
Punjab Kings                                                         -0.237365
Rajasthan Royals                                                      0.650200
Rising Pune Supergiant                                               -0.087159
Royal Challengers Bangalore                                           0.481890
Sunrisers Hyderabad                                                   0.101432
```

```
nvenue                       Brabourne Stadium  Buffalo Park  \
team
Chennai Super Kings                   0.484718      0.135386
Deccan Chargers                      -0.030286      0.323066
Delhi Capitals                        0.197712     -0.219705
Gujarat Lions                        -0.234550     -0.856506
Gujarat Titans                        0.356644     -0.624410
Kochi Tuskers Kerala                 -0.510807     -0.244046
Kolkata Knight Riders                -0.067700     -0.644347
Lucknow Super Giants                  0.635285      0.124383
Mumbai Indians                        0.413761     -0.324148
Pune Warriors                        -0.315053      0.414876
Punjab Kings                          0.274853     -0.811602
Rajasthan Royals                      0.016968     -0.421080
Rising Pune Supergiant               -0.241387     -0.458487
Royal Challengers Bangalore           0.298452     -0.403164
Sunrisers Hyderabad                   0.246625     -0.204142

nvenue                       De Beers Diamond Oval  \
team
Chennai Super Kings                      -0.012121
Deccan Chargers                          -0.091315
Delhi Capitals                           -0.582670
Gujarat Lions                            -0.376351
Gujarat Titans                           -0.380534
Kochi Tuskers Kerala                     -0.025598
Kolkata Knight Riders                    -0.477656
Lucknow Super Giants                     -0.232032
Mumbai Indians                           -0.582548
Pune Warriors                            -0.044923
Punjab Kings                             -0.680628
Rajasthan Royals                         -0.156811
Rising Pune Supergiant                   -0.294295
Royal Challengers Bangalore              -0.468081
Sunrisers Hyderabad                      -0.418742

nvenue                       Dr DY Patil Sports Academy  \
team
Chennai Super Kings                            0.064242
Deccan Chargers                               -0.191570
Delhi Capitals                                -0.385787
Gujarat Lions                                 -0.029822
Gujarat Titans                                 0.332866
Kochi Tuskers Kerala                          -0.125408
Kolkata Knight Riders                         -0.036331
Lucknow Super Giants                           0.191303
Mumbai Indians                                 0.080934
```

```
Pune Warriors                                -0.460757
Punjab Kings                                  0.059835
Rajasthan Royals                             -0.296194
Rising Pune Supergiant                       -0.067299
Royal Challengers Bangalore                  -0.045532
Sunrisers Hyderabad                           0.091481

nvenue                       Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium
\
team
Chennai Super Kings                                                   -0.008847
Deccan Chargers                                                        0.309281
Delhi Capitals                                                        -0.235605
Gujarat Lions                                                         -0.592285
Gujarat Titans                                                        -0.200639
Kochi Tuskers Kerala                                                  -0.392605
Kolkata Knight Riders                                                 -0.667569
Lucknow Super Giants                                                   0.094114
Mumbai Indians                                                        -0.258875
Pune Warriors                                                          0.112682
Punjab Kings                                                          -0.446720
Rajasthan Royals                                                      -0.654914
Rising Pune Supergiant                                                -0.433444
Royal Challengers Bangalore                                           -0.020011
Sunrisers Hyderabad                                                   -0.358764

nvenue                       Dubai International Cricket Stadium  …  \
team                                                                 …
Chennai Super Kings                                    -0.704793  …
Deccan Chargers                                        -0.488496  …
Delhi Capitals                                         -0.065422  …
Gujarat Lions                                           0.037752  …
Gujarat Titans                                         -0.494966  …
Kochi Tuskers Kerala                                   -0.524344  …
Kolkata Knight Riders                                  -0.175180  …
Lucknow Super Giants                                   -0.204802  …
Mumbai Indians                                         -0.159297  …
Pune Warriors                                          -0.407698  …
Punjab Kings                                           -0.026712  …
Rajasthan Royals                                       -0.258076  …
Rising Pune Supergiant                                 -0.581642  …
Royal Challengers Bangalore                            -0.152400  …
Sunrisers Hyderabad                                     0.206767  …

nvenue                       Sawai Mansingh Stadium  \
team
Chennai Super Kings                        0.614295
```

```
Deccan Chargers                    0.072525
Delhi Capitals                     0.251447
Gujarat Lions                      0.272745
Gujarat Titans                     0.334150
Kochi Tuskers Kerala               0.550372
Kolkata Knight Riders              0.669986
Lucknow Super Giants               0.066734
Mumbai Indians                     0.484998
Pune Warriors                      0.186463
Punjab Kings                       0.220112
Rajasthan Royals                   0.581346
Rising Pune Supergiant             0.547560
Royal Challengers Bangalore        0.056850
Sunrisers Hyderabad                0.287506

nvenue                       Shaheed Veer Narayan Singh International Stadium  \
team
Chennai Super Kings                                                 -0.330481
Deccan Chargers                                                     -0.145522
Delhi Capitals                                                      -0.690510
Gujarat Lions                                                       -0.148743
Gujarat Titans                                                      -0.377590
Kochi Tuskers Kerala                                                 0.312969
Kolkata Knight Riders                                               -0.260862
Lucknow Super Giants                                                -0.396378
Mumbai Indians                                                      -0.700679
Pune Warriors                                                       -0.147146
Punjab Kings                                                        -0.493278
Rajasthan Royals                                                    -0.179927
Rising Pune Supergiant                                               0.072109
Royal Challengers Bangalore                                         -0.633312
Sunrisers Hyderabad                                                 -0.418352

nvenue                       Sharjah Cricket Stadium  Sheikh Zayed Stadium  \
team
Chennai Super Kings                        -0.344795             -0.678389
Deccan Chargers                            -0.051093             -0.335507
Delhi Capitals                             -0.672109             -0.245990
Gujarat Lions                               0.272769              0.181267
Gujarat Titans                              0.316139             -0.357537
Kochi Tuskers Kerala                        0.077981             -0.344323
Kolkata Knight Riders                      -0.280869             -0.168458
Lucknow Super Giants                       -0.507764             -0.469784
Mumbai Indians                             -0.533884             -0.282940
Pune Warriors                              -0.419952             -0.301183
Punjab Kings                               -0.016475             -0.074524
Rajasthan Royals                           -0.384506             -0.246592
```

```
Rising Pune Supergiant                     0.023772             -0.493513
Royal Challengers Bangalore               -0.096105             -0.290649
Sunrisers Hyderabad                       -0.583865              0.091420

nvenue                       St George's Park  Subrata Roy Sahara Stadium  \
team
Chennai Super Kings                  0.175970                    0.371378
Deccan Chargers                      0.390802                    0.846747
Delhi Capitals                      -0.440041                    0.012019
Gujarat Lions                       -0.402363                   -0.337456
Gujarat Titans                      -0.127552                   -0.025156
Kochi Tuskers Kerala                -0.181523                    0.193316
Kolkata Knight Riders               -0.587800                   -0.194175
Lucknow Super Giants                -0.253298                   -0.153607
Mumbai Indians                      -0.393057                   -0.030536
Pune Warriors                        0.291235                    0.732204
Punjab Kings                        -0.606187                   -0.297540
Rajasthan Royals                    -0.340468                   -0.203694
Rising Pune Supergiant              -0.425234                    0.069797
Royal Challengers Bangalore         -0.351099                   -0.269132
Sunrisers Hyderabad                 -0.354035                   -0.134715

nvenue                       SuperSport Park  \
team
Chennai Super Kings                -0.290208
Deccan Chargers                    -0.006110
Delhi Capitals                     -0.887406
Gujarat Lions                      -0.014408
Gujarat Titans                     -0.076099
Kochi Tuskers Kerala                0.208852
Kolkata Knight Riders              -0.406551
Lucknow Super Giants               -0.579750
Mumbai Indians                     -0.797763
Pune Warriors                      -0.220594
Punjab Kings                       -0.434986
Rajasthan Royals                   -0.293098
Rising Pune Supergiant             -0.036324
Royal Challengers Bangalore        -0.617185
Sunrisers Hyderabad                -0.576295

nvenue                       Vidarbha Cricket Association Stadium  \
team
Chennai Super Kings                                     -0.091969
Deccan Chargers                                          0.320812
Delhi Capitals                                          -0.347624
Gujarat Lions                                           -0.782535
Gujarat Titans                                          -0.594760
```

```
Kochi Tuskers Kerala                                     -0.133825
Kolkata Knight Riders                                    -0.590540
Lucknow Super Giants                                      0.107329
Mumbai Indians                                           -0.446074
Pune Warriors                                             0.253771
Punjab Kings                                             -0.630984
Rajasthan Royals                                         -0.502825
Rising Pune Supergiant                                   -0.269292
Royal Challengers Bangalore                              -0.602193
Sunrisers Hyderabad                                      -0.175590

nvenue                       Wankhede Stadium  Zayed Cricket Stadium
team
Chennai Super Kings                  0.458742              -0.321804
Deccan Chargers                     -0.089658               0.071895
Delhi Capitals                       0.444943              -0.805102
Gujarat Lions                        0.484369               0.152425
Gujarat Titans                       0.751623               0.245865
Kochi Tuskers Kerala                 0.030634               0.071750
Kolkata Knight Riders                0.703936              -0.405405
Lucknow Super Giants                 0.360653              -0.557175
Mumbai Indians                       0.817739              -0.661220
Pune Warriors                       -0.312497              -0.336383
Punjab Kings                         0.828338              -0.152552
Rajasthan Royals                     0.391604              -0.441228
Rising Pune Supergiant               0.343401              -0.039657
Royal Challengers Bangalore          0.457169              -0.317032
Sunrisers Hyderabad                  0.558477              -0.610111

[15 rows x 40 columns]
```

[47]:
```python
#get recommendations for teams:
def recommend_venue(team , smat):
    return smat.loc[team].sort_values(ascending = False)[:5].index.tolist()

nteams = sorted(df['team1'].unique())

for team in nteams:
    print(f"{team}: {recommend_venue(team , smat)}")
```

```
Chennai Super Kings: ['Rajiv Gandhi International Stadium', 'Sawai Mansingh
Stadium', 'Eden Gardens', 'Punjab Cricket Association IS Bindra Stadium',
'Brabourne Stadium']
Deccan Chargers: ['Subrata Roy Sahara Stadium', 'Nehru Stadium', 'Feroz Shah
Kotla', 'Kingsmead', "St George's Park"]
Delhi Capitals: ['M.Chinnaswamy Stadium', 'Feroz Shah Kotla', 'Arun Jaitley
Stadium', 'Punjab Cricket Association IS Bindra Stadium', 'Bharat Ratna Shri
Atal Bihari Vajpayee Ekana Cricket Stadium']
```

```
Gujarat Lions: ['Saurashtra Cricket Association Stadium', 'Green Park',
'Narendra Modi Stadium', 'Holkar Cricket Stadium', 'Punjab Cricket Association
IS Bindra Stadium']
Gujarat Titans: ['Narendra Modi Stadium', 'Wankhede Stadium', 'Punjab Cricket
Association IS Bindra Stadium', 'Maharaja Yadavindra Singh International Cricket
Stadium', 'Eden Gardens']
Kochi Tuskers Kerala: ['Sawai Mansingh Stadium', 'Eden Gardens', 'Feroz Shah
Kotla', 'Punjab Cricket Association IS Bindra Stadium', 'Nehru Stadium']
Kolkata Knight Riders: ['Eden Gardens', 'Punjab Cricket Association IS Bindra
Stadium', 'Wankhede Stadium', 'Rajiv Gandhi International Stadium', 'Sawai
Mansingh Stadium']
Lucknow Super Giants: ['Brabourne Stadium', 'Bharat Ratna Shri Atal Bihari
Vajpayee Ekana Cricket Stadium', 'Arun Jaitley Stadium', 'Barsapara Cricket
Stadium', 'Himachal Pradesh Cricket Association Stadium']
Mumbai Indians: ['Arun Jaitley Stadium', 'Wankhede Stadium', 'Rajiv Gandhi
International Stadium', 'Punjab Cricket Association IS Bindra Stadium', 'Eden
Gardens']
Pune Warriors: ['Feroz Shah Kotla', 'Subrata Roy Sahara Stadium', 'Nehru
Stadium', 'Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Stadium',
'Buffalo Park']
Punjab Kings: ['Wankhede Stadium', 'Punjab Cricket Association IS Bindra
Stadium', 'Narendra Modi Stadium', 'Saurashtra Cricket Association Stadium',
'Eden Gardens']
Rajasthan Royals: ['Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket
Stadium', 'M.Chinnaswamy Stadium', 'Sawai Mansingh Stadium', 'Eden Gardens',
'Feroz Shah Kotla']
Rising Pune Supergiant: ['Punjab Cricket Association IS Bindra Stadium', 'Eden
Gardens', 'Feroz Shah Kotla', 'Sawai Mansingh Stadium', 'Nehru Stadium']
Royal Challengers Bangalore: ['M.Chinnaswamy Stadium', 'Maharashtra Cricket
Association Stadium', 'Himachal Pradesh Cricket Association Stadium', 'Feroz
Shah Kotla', 'Bharat Ratna Shri Atal Bihari Vajpayee Ekana Cricket Stadium']
Sunrisers Hyderabad: ['Arun Jaitley Stadium', 'Rajiv Gandhi International
Stadium', 'Maharaja Yadavindra Singh International Cricket Stadium', 'MA
Chidambaram Stadium', 'Wankhede Stadium']
```

[ ]:

[ ]:

[ ]:

[ ]:

7 Question 7: Performance Trend Analysis (8 points)

7.0.1 Task:

Use **Linear Regression** to analyze performance trends:

1. Track team performance over seasons:
    - Create yearly win percentage for each team
    - Fit linear regression to identify improving/declining teams
2. Predict next season performance
3. Identify factors affecting performance trends:
    - Player retention
    - Home ground advantage
    - Toss luck factor

7.0.2 Visualization:

- Time series plot with regression lines for top 5 teams

[48]:
```
#yearly win percentage for each team
cols = ['team1' , 'team2' , 'season' , 'win']
# df[cols]

# convert (t, t, s, w) -> (t , s, w):
cdf = pd.concat([df[['team1'] + cols[2:]].rename(columns = {'team1':'team'}) ,␣
↪df[['team2'] + cols[2:]].rename(columns = {'team2' : 'team'})])
gdf = cdf.groupby(['team' , 'season']).count().unstack(level = 1).fillna(0)
gdf.columns = [col[1] if isinstance(col , tuple) else col[0] for col in gdf.
↪columns]
g2df = cdf.groupby('season')['win'].count()
```

[49]:
```
for col in gdf.columns:
    gdf[col] = (gdf[col] / g2df.loc[col]) * 100
gdf
```

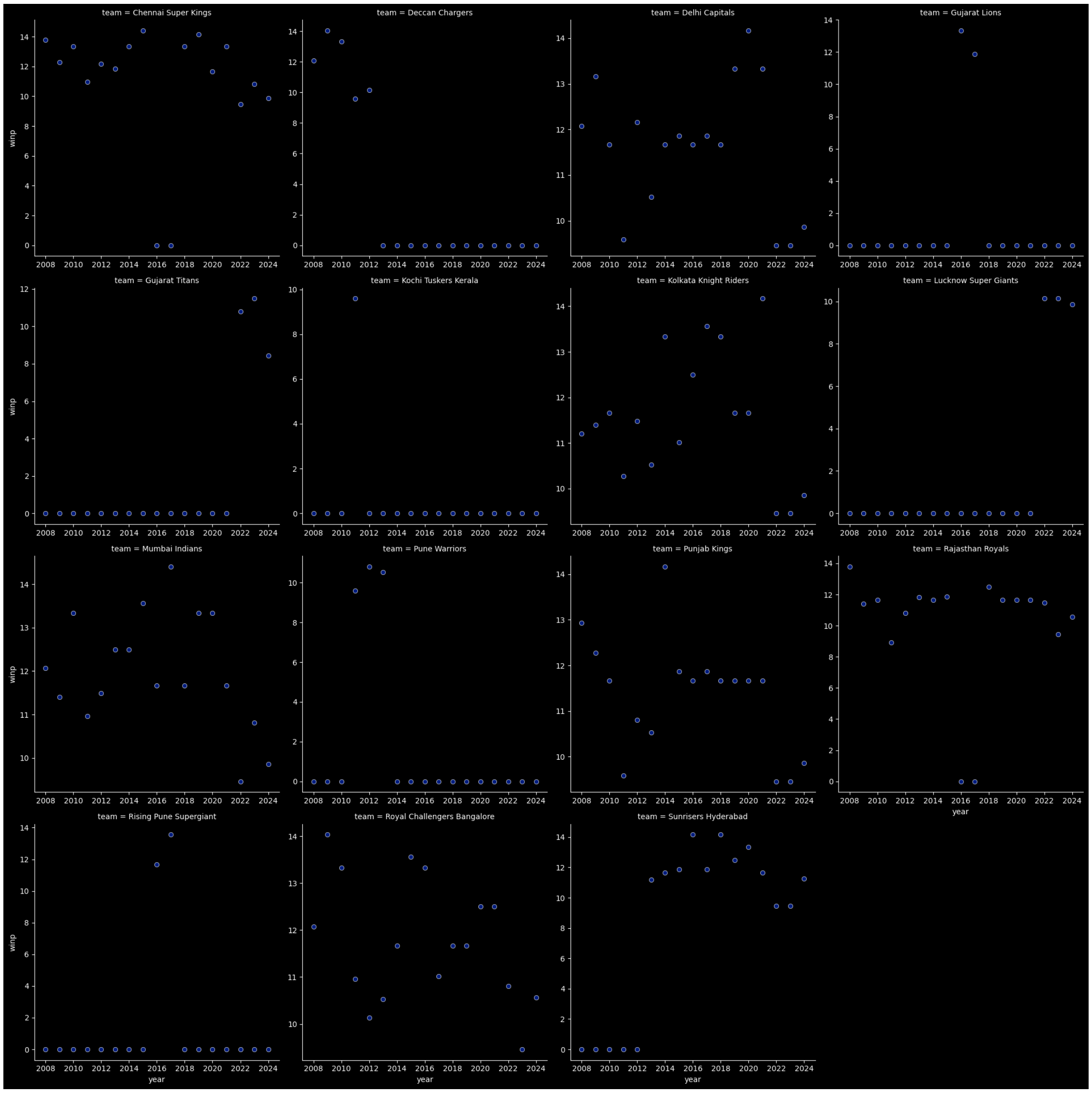
[49]:
```
                                  2008       2009       2010       2011  \
team
Chennai Super Kings          13.793103  12.280702  13.333333  10.958904
Deccan Chargers              12.068966  14.035088  13.333333   9.589041
Delhi Capitals               12.068966  13.157895  11.666667   9.589041
Gujarat Lions                 0.000000   0.000000   0.000000   0.000000
Gujarat Titans                0.000000   0.000000   0.000000   0.000000
Kochi Tuskers Kerala          0.000000   0.000000   0.000000   9.589041
Kolkata Knight Riders        11.206897  11.403509  11.666667  10.273973
Lucknow Super Giants          0.000000   0.000000   0.000000   0.000000
Mumbai Indians               12.068966  11.403509  13.333333  10.958904
Pune Warriors                 0.000000   0.000000   0.000000   9.589041
Punjab Kings                 12.931034  12.280702  11.666667   9.589041
Rajasthan Royals             13.793103  11.403509  11.666667   8.904110
Rising Pune Supergiant        0.000000   0.000000   0.000000   0.000000
Royal Challengers Bangalore  12.068966  14.035088  13.333333  10.958904
Sunrisers Hyderabad           0.000000   0.000000   0.000000   0.000000
```

```
                                  2012       2013       2014       2015  \
team
Chennai Super Kings          12.162162  11.842105  13.333333  14.406780
Deccan Chargers              10.135135   0.000000   0.000000   0.000000
Delhi Capitals               12.162162  10.526316  11.666667  11.864407
Gujarat Lions                 0.000000   0.000000   0.000000   0.000000
Gujarat Titans                0.000000   0.000000   0.000000   0.000000
Kochi Tuskers Kerala          0.000000   0.000000   0.000000   0.000000
Kolkata Knight Riders        11.486486  10.526316  13.333333  11.016949
Lucknow Super Giants          0.000000   0.000000   0.000000   0.000000
Mumbai Indians               11.486486  12.500000  12.500000  13.559322
Pune Warriors                10.810811  10.526316   0.000000   0.000000
Punjab Kings                 10.810811  10.526316  14.166667  11.864407
Rajasthan Royals             10.810811  11.842105  11.666667  11.864407
Rising Pune Supergiant        0.000000   0.000000   0.000000   0.000000
Royal Challengers Bangalore  10.135135  10.526316  11.666667  13.559322
Sunrisers Hyderabad           0.000000  11.184211  11.666667  11.864407

                                  2016       2017       2018       2019  \
team
Chennai Super Kings           0.000000   0.000000  13.333333  14.166667
Deccan Chargers               0.000000   0.000000   0.000000   0.000000
Delhi Capitals               11.666667  11.864407  11.666667  13.333333
Gujarat Lions                13.333333  11.864407   0.000000   0.000000
Gujarat Titans                0.000000   0.000000   0.000000   0.000000
Kochi Tuskers Kerala          0.000000   0.000000   0.000000   0.000000
Kolkata Knight Riders        12.500000  13.559322  13.333333  11.666667
Lucknow Super Giants          0.000000   0.000000   0.000000   0.000000
Mumbai Indians               11.666667  14.406780  11.666667  13.333333
Pune Warriors                 0.000000   0.000000   0.000000   0.000000
Punjab Kings                 11.666667  11.864407  11.666667  11.666667
Rajasthan Royals              0.000000   0.000000  12.500000  11.666667
Rising Pune Supergiant       11.666667  13.559322   0.000000   0.000000
Royal Challengers Bangalore  13.333333  11.016949  11.666667  11.666667
Sunrisers Hyderabad          14.166667  11.864407  14.166667  12.500000

                                  2020       2021       2022       2023  \
team
Chennai Super Kings          11.666667  13.333333   9.459459  10.810811
Deccan Chargers               0.000000   0.000000   0.000000   0.000000
Delhi Capitals               14.166667  13.333333   9.459459   9.459459
Gujarat Lions                 0.000000   0.000000   0.000000   0.000000
Gujarat Titans                0.000000   0.000000  10.810811  11.486486
Kochi Tuskers Kerala          0.000000   0.000000   0.000000   0.000000
Kolkata Knight Riders        11.666667  14.166667   9.459459   9.459459
Lucknow Super Giants          0.000000   0.000000  10.135135  10.135135
```

```
Mumbai Indians               13.333333  11.666667   9.459459  10.810811
Pune Warriors                 0.000000   0.000000   0.000000   0.000000
Punjab Kings                 11.666667  11.666667   9.459459   9.459459
Rajasthan Royals             11.666667  11.666667  11.486486   9.459459
Rising Pune Supergiant        0.000000   0.000000   0.000000   0.000000
Royal Challengers Bangalore  12.500000  12.500000  10.810811   9.459459
Sunrisers Hyderabad          13.333333  11.666667   9.459459   9.459459

                                  2024
team
Chennai Super Kings           9.859155
Deccan Chargers               0.000000
Delhi Capitals                9.859155
Gujarat Lions                 0.000000
Gujarat Titans                8.450704
Kochi Tuskers Kerala          0.000000
Kolkata Knight Riders         9.859155
Lucknow Super Giants          9.859155
Mumbai Indians                9.859155
Pune Warriors                 0.000000
Punjab Kings                  9.859155
Rajasthan Royals             10.563380
Rising Pune Supergiant        0.000000
Royal Challengers Bangalore  10.563380
Sunrisers Hyderabad          11.267606
```

[50]:
```
mdf = gdf.reset_index().melt(id_vars = 'team' , var_name = 'year' , value_name='winp')
plt.figure(figsize = (20 , 20))
sns.relplot(data = mdf , x = 'year' , y = 'winp' , col = 'team' , col_wrap = 4, kind = 'scatter' , facet_kws= {'sharey' : False , 'sharex' : False})
plt.savefig('./data/q7_winpbyyear.png')
```

```
<Figure size 2000x2000 with 0 Axes>
```



[ ]:

[ ]:

[ ]:

8 Question 8: Player of Match Prediction - KNN (8 points)

8.0.1 Task:

Use **K-Nearest Neighbors (KNN)** to predict potential 'Player of Match':

1. Create match context features:

- Venue characteristics
- Team strengths
- Historical player performance at venue

2. Use KNN to find similar historical matches
3. Predict likely ‘Player of Match’ based on similar matches
4. Optimize K using cross-validation

[59]:
```
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, classification_report,␣
↪confusion_matrix

# Label encode categorical columns
label_cols = ['nvenue', 'team1', 'team2', 'toss_winner', 'winner',␣
↪'player_of_match']
tdf = df[label_cols].copy()
encoders = {col: LabelEncoder() for col in label_cols}
for col in label_cols:
    enc = LabelEncoder()
    tdf[col] = enc.fit_transform(tdf[col])
    encoders[col] = enc

# Aggregate player historical performance
player_perf = tdf.groupby('player_of_match')['winner'].count().reset_index()
player_perf.columns = ['player_of_match', 'match_wins']
tdf = tdf.merge(player_perf, on='player_of_match', how='left')

# Create context features
tdf['venue_strength'] = tdf.groupby('nvenue')['winner'].transform(lambda x: x.
↪value_counts().max())
tdf['team_strength_diff'] = abs(tdf['team1'] - tdf['team2'])

# Final feature set
features = ['nvenue', 'team1', 'team2', 'venue_strength', 'team_strength_diff']
target = 'player_of_match'

x = tdf[features]
y = tdf[target]

xtrain , xtest , ytrain , ytest = train_test_split(x , y , test_size= 0.2 ,␣
↪random_state= 42)
```

[60]:
```
# Parameter tuning for best K
param_grid = {'n_neighbors': list(range(3, 21))}
grid = GridSearchCV(KNeighborsClassifier(), param_grid, cv=5,␣
↪scoring='accuracy')
grid.fit(xtrain, ytrain)

best_k = grid.best_params_['n_neighbors']
best_score = grid.best_score_
print(f"  Best K = {best_k} with CV Accuracy = {best_score:.3f}")
```

```
/opt/anaconda3/envs/tensorflow/lib/python3.11/site-
packages/sklearn/model_selection/_split.py:805: UserWarning: The least populated
class in y has only 1 members, which is less than n_splits=5.
  warnings.warn(
```

```
 Best K = 17 with CV Accuracy = 0.046
```

[62]:
```
best_knn = KNeighborsClassifier(n_neighbors=best_k)
best_knn.fit(xtrain, ytrain)
y_pred_best = best_knn.predict(xtest)
print("Final Accuracy:", accuracy_score(ytest, y_pred_best))
```

```
Final Accuracy: 0.0319634703196347
```

[73]:
```
sample = pd.DataFrame({
    'nvenue': [encoders['nvenue'].transform(['Eden Gardens'])[0]],
    'team1': [encoders['team1'].transform(['Mumbai Indians'])[0]],
    'team2': [encoders['team2'].transform(['Chennai Super Kings'])[0]],
    'venue_strength': [25],
    'team_strength_diff': [abs(5 - 12)]
})
pred_player = encoders['player_of_match'].inverse_transform(best_knn.
↪predict(sample))
print(f"Predicted Player of the Match: {pred_player[0]}")
```

```
Predicted Player of the Match: A Mishra
```

9 Question 9: Toss Decision Strategy - Decision Tree (9 points)

9.0.1 Task:

Build a **Decision Tree** to recommend toss decisions:

1. Create a model to predict optimal toss decision (bat/field) based on:
    - Venue history
    - Weather conditions (create synthetic if not available)
    - Team strengths
    - Match importance
2. Visualize the decision tree (max_depth=5)

3. Extract decision rules in plain English
4. Calculate feature importance

9.0.2 Business Application:

"Provide captains with data-driven toss decision recommendations"

[51]:
```
from sklearn.tree import DecisionTreeClassifier , plot_tree , export_text
from sklearn.preprocessing import OrdinalEncoder , LabelEncoder

#compute toss-strength for each team:
teams = pd.concat([df['team1'] , df['team2']]).unique()
thist = {t : {'played' : 0 , 'won' : 0}for t in teams}
#iterate over the dataframe to compute stats prior to each match:
toss_win_strength = []
opp_strength = []
for id , row in df.iterrows():
    tw = row['toss_winner']
    if tw == row['team1']: opp = row['team2']
    else: opp = row['team1']

    #get winrate so far:
    def get_wr(team):
        if team not in thist:
            return 0.5
        played = thist[team]['played']
        return thist[team]['won'] / played if played > 0 else 0.5

    tws = get_wr(tw)
    opps = get_wr(opp)

    toss_win_strength.append(tws)
    opp_strength.append(opps)

    winner = row['winner']

    #update history:
    for t in [row['team1'] , row['team2']]:
        thist[t]['played'] += 1

    thist.setdefault(winner , {'played' : 0 , 'won' : 0})
    thist[winner]['won'] +=1


df['toss_winner_strength'] = toss_win_strength
df['opp_strength'] = opp_strength
```

```
cols = ['date' , 'nvenue' , 'team1' , 'team2' , 'toss_winner_strength' ,␣
↪'opp_strength']
ddf = df[cols].copy()

#translate string to numeric features:
ddf['match_importance_code'] = pd.Categorical(df['match_importance'] ,␣
↪categories= df['match_importance'].cat.categories.to_list()).codes
y = df['toss_winner']
le = LabelEncoder()
ddf['evenue'] = le.fit_transform(ddf['nvenue'])
le.fit(pd.concat([ddf['team1'] , ddf['team2']]).unique())
ddf['eteam1'] = le.transform(ddf['team1'])
ddf['eteam2'] = le.transform(ddf['team2'])
ddf = ddf.iloc[: , 4:]
ddf
```

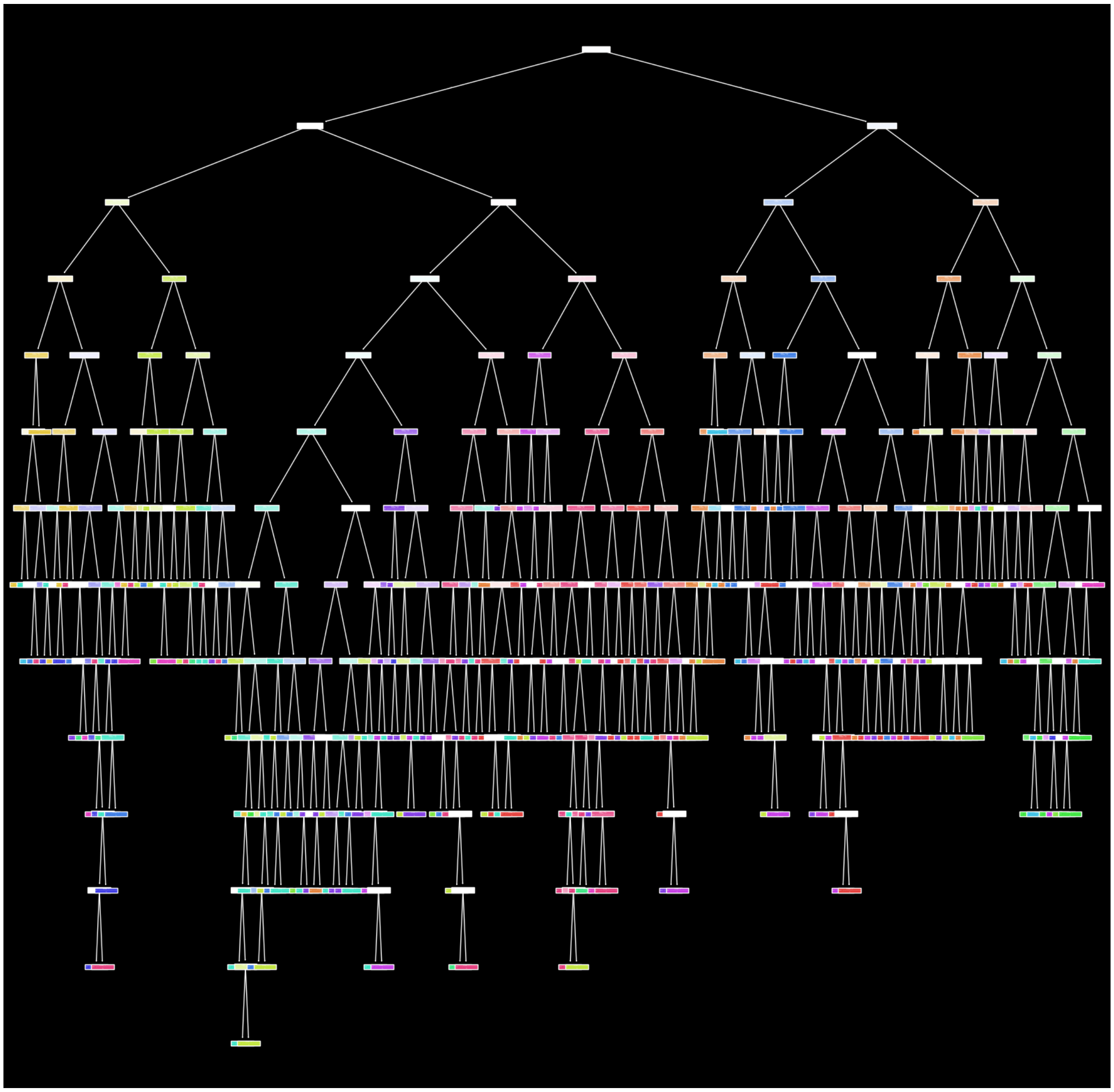
[51]:

```
      toss_winner_strength  opp_strength  match_importance_code  evenue  \
0                 0.500000      0.500000                      0      17
1                 0.500000      0.500000                      0      26
2                 0.500000      0.500000                      0      11
3                 0.500000      0.000000                      0      38
4                 0.500000      1.000000                      0      10
…                        …             …                    …       …
1090              0.457143      0.483146                      0      27
1091              0.486034      0.518072                      1      21
1092              0.506849      0.484252                      1      21
1093              0.509091      0.483333                      1      18
1094              0.486188      0.520000                      2      18

      eteam1  eteam2
0         13       6
1         10       0
2          2      11
3          8      13
4          6       1
…      …      …
1090      10      14
1091      14       6
1092      13      11
1093      14      11
1094      14       6

[1095 rows x 6 columns]
```

[52]:

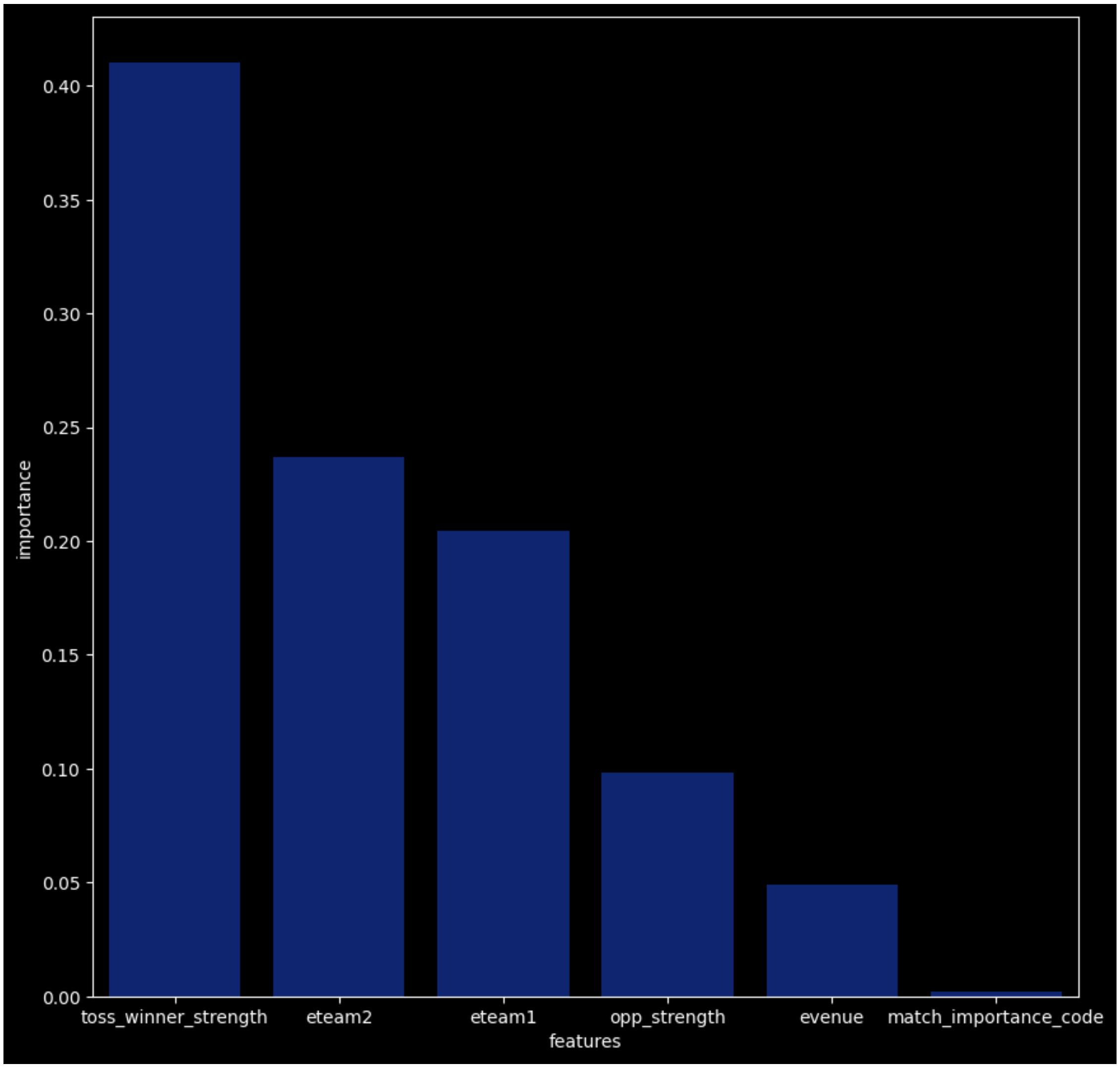
```
xtrain , xtest , ytrain , ytest = train_test_split(ddf , y , test_size= 0.2 ,␣
↪random_state = 42)
model = DecisionTreeClassifier(criterion = 'entropy')
model.fit(xtrain , ytrain)
print(f"model score: {model.score(xtest , ytest)}")
plot_tree(model , filled = True)
plt.show()
```

```
model score: 0.6301369863013698
```



[53]:
```
tdf = pd.DataFrame({'features' : xtrain.columns.tolist() , 'importance' : model.
↪feature_importances_}).sort_values(by = 'importance' , ascending = False)
plt.figure(figsize = (10 , 10))
```

```python
sns.barplot(tdf , x = 'features' , y = 'importance')
plt.show()
```



[54]:

```python
#decision rules in english:
rules = export_text(model , feature_names= xtrain.columns.tolist())
print(rules)
```

```
|--- toss_winner_strength <= 0.54
|   |--- toss_winner_strength <= 0.46
|   |   |--- toss_winner_strength <= 0.41
|   |   |   |--- eteam1 <= 1.50
|   |   |   |   |--- eteam1 <= 0.50
|   |   |   |   |   |--- eteam2 <= 6.50
|   |   |   |   |   |   |--- eteam2 <= 3.50
|   |   |   |   |   |   |   |--- class: Deccan Chargers
```

```
|   |   |   |   |   |   |--- eteam2 >  3.50
|   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |--- eteam2 >  6.50
|   |   |   |   |   |   |--- eteam2 <= 8.50
|   |   |   |   |   |   |   |--- evenue <= 11.00
|   |   |   |   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |   |   |   |   |--- evenue >  11.00
|   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |--- eteam2 >  8.50
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.22
|   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.22
|   |   |   |   |   |   |   |   |--- class: Pune Warriors
|   |   |   |   |--- eteam1 >  0.50
|   |   |   |   |   |--- class: Deccan Chargers
|   |   |   |--- eteam1 >  1.50
|   |   |   |   |--- eteam2 <= 1.50
|   |   |   |   |   |--- evenue <= 10.50
|   |   |   |   |   |   |--- eteam1 <= 7.50
|   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |--- eteam1 >  7.50
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.40
|   |   |   |   |   |   |   |   |--- class: Deccan Chargers
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.40
|   |   |   |   |   |   |   |   |--- class: Pune Warriors
|   |   |   |   |   |--- evenue >  10.50
|   |   |   |   |   |   |--- toss_winner_strength <= 0.41
|   |   |   |   |   |   |   |--- class: Deccan Chargers
|   |   |   |   |   |   |--- toss_winner_strength >  0.41
|   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |--- eteam2 >  1.50
|   |   |   |   |   |--- toss_winner_strength <= 0.31
|   |   |   |   |   |   |--- toss_winner_strength <= 0.10
|   |   |   |   |   |   |   |--- evenue <= 28.00
|   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |   |--- evenue >  28.00
|   |   |   |   |   |   |   |   |--- evenue <= 34.00
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |--- evenue >  34.00
|   |   |   |   |   |   |   |   |   |--- class: Kochi Tuskers Kerala
|   |   |   |   |   |   |--- toss_winner_strength >  0.10
|   |   |   |   |   |   |   |--- opp_strength <= 0.61
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.45
|   |   |   |   |   |   |   |   |   |--- class: Rising Pune Supergiant
|   |   |   |   |   |   |   |   |--- opp_strength >  0.45
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.22
|   |   |   |   |   |   |   |   |   |   |--- class: Rising Pune Supergiant
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.22
```

```
|   |   |   |   |   |   |   |   |   |   |--- eteam2 <= 5.00
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- eteam2 >  5.00
|   |   |   |   |   |   |   |   |   |   |   |--- class: Pune Warriors
|   |   |   |   |   |   |   |--- opp_strength >  0.61
|   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |--- toss_winner_strength >  0.31
|   |   |   |   |   |   |--- eteam1 <= 11.50
|   |   |   |   |   |   |   |--- evenue <= 34.50
|   |   |   |   |   |   |   |   |--- eteam1 <= 4.00
|   |   |   |   |   |   |   |   |   |--- class: Kochi Tuskers Kerala
|   |   |   |   |   |   |   |   |--- eteam1 >  4.00
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.76
|   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.76
|   |   |   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |   |--- evenue >  34.50
|   |   |   |   |   |   |   |   |--- class: Pune Warriors
|   |   |   |   |   |   |--- eteam1 >  11.50
|   |   |   |   |   |   |   |--- evenue <= 18.50
|   |   |   |   |   |   |   |   |--- class: Pune Warriors
|   |   |   |   |   |   |   |--- evenue >  18.50
|   |   |   |   |   |   |   |   |--- class: Rising Pune Supergiant
|   |   |--- toss_winner_strength >  0.41
|   |   |   |--- eteam2 <= 4.50
|   |   |   |   |--- eteam2 <= 1.50
|   |   |   |   |   |--- opp_strength <= 0.59
|   |   |   |   |   |   |--- match_importance_code <= 1.50
|   |   |   |   |   |   |   |--- class: Deccan Chargers
|   |   |   |   |   |   |--- match_importance_code >  1.50
|   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |--- opp_strength >  0.59
|   |   |   |   |   |   |--- evenue <= 23.00
|   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |--- evenue >  23.00
|   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |--- eteam2 >  1.50
|   |   |   |   |   |--- eteam2 <= 2.50
|   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |--- eteam2 >  2.50
|   |   |   |   |   |   |--- eteam1 <= 6.00
|   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |--- eteam1 >  6.00
|   |   |   |   |   |   |   |--- eteam1 <= 11.00
|   |   |   |   |   |   |   |   |--- class: Gujarat Lions
|   |   |   |   |   |   |   |--- eteam1 >  11.00
|   |   |   |   |   |   |   |   |--- class: Rising Pune Supergiant
|   |   |   |--- eteam2 >  4.50
```

```
|   |   |   |   |--- eteam1 <= 3.50
|   |   |   |   |   |--- eteam1 <= 1.50
|   |   |   |   |   |   |--- eteam2 <= 7.00
|   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |--- eteam2 >  7.00
|   |   |   |   |   |   |   |--- class: Deccan Chargers
|   |   |   |   |   |--- eteam1 >  1.50
|   |   |   |   |   |   |--- toss_winner_strength <= 0.45
|   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |--- toss_winner_strength >  0.45
|   |   |   |   |   |   |   |--- opp_strength <= 0.53
|   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |--- opp_strength >  0.53
|   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |--- eteam1 >  3.50
|   |   |   |   |   |--- toss_winner_strength <= 0.44
|   |   |   |   |   |   |--- opp_strength <= 0.60
|   |   |   |   |   |   |   |--- eteam1 <= 5.50
|   |   |   |   |   |   |   |   |--- class: Kochi Tuskers Kerala
|   |   |   |   |   |   |   |--- eteam1 >  5.50
|   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |--- opp_strength >  0.60
|   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |--- toss_winner_strength >  0.44
|   |   |   |   |   |   |--- eteam1 <= 7.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.51
|   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |--- opp_strength >  0.51
|   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |--- eteam1 >  7.00
|   |   |   |   |   |   |   |--- eteam2 <= 7.00
|   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |   |--- eteam2 >  7.00
|   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |--- toss_winner_strength >  0.46
|   |   |--- eteam2 <= 10.50
|   |   |   |--- eteam1 <= 12.50
|   |   |   |   |--- eteam2 <= 8.50
|   |   |   |   |   |--- eteam1 <= 9.50
|   |   |   |   |   |   |--- eteam2 <= 5.00
|   |   |   |   |   |   |   |--- eteam1 <= 5.50
|   |   |   |   |   |   |   |   |--- eteam1 <= 4.50
|   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |--- eteam1 >  4.50
|   |   |   |   |   |   |   |   |   |--- class: Kochi Tuskers Kerala
|   |   |   |   |   |   |   |--- eteam1 >  5.50
|   |   |   |   |   |   |   |   |--- eteam1 <= 6.50
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.81
```

```
|   |   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.49
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.49
|   |   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.81
|   |   |   |   |   |   |   |   |   |   |--- class: Deccan Chargers
|   |   |   |   |   |   |   |   |--- eteam1 >  6.50
|   |   |   |   |   |   |   |   |   |--- eteam1 <= 7.50
|   |   |   |   |   |   |   |   |   |   |--- class: Gujarat Titans
|   |   |   |   |   |   |   |   |   |--- eteam1 >  7.50
|   |   |   |   |   |   |   |   |   |   |--- evenue <= 5.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |   |--- evenue >  5.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |--- eteam2 >  5.00
|   |   |   |   |   |   |   |--- eteam2 <= 7.00
|   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.53
|   |   |   |   |   |   |   |   |   |--- evenue <= 36.50
|   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |   |--- evenue >  36.50
|   |   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.56
|   |   |   |   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |   |   |   |   |--- opp_strength >  0.56
|   |   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.53
|   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |--- eteam2 >  7.00
|   |   |   |   |   |   |   |   |--- eteam1 <= 2.50
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.49
|   |   |   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.49
|   |   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |--- eteam1 >  2.50
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.50
|   |   |   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.50
|   |   |   |   |   |   |   |   |   |   |--- eteam1 <= 4.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Gujarat Lions
|   |   |   |   |   |   |   |   |   |   |--- eteam1 >  4.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |--- eteam1 >  9.50
|   |   |   |   |   |   |--- eteam1 <= 10.50
|   |   |   |   |   |   |   |--- eteam2 <= 5.00
|   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.50
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.50
|   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.50
|   |   |   |   |   |   |   |   |   |   |--- evenue <= 19.50
```

```
|   |   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |   |   |--- evenue >  19.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.50
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.45
|   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.45
|   |   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |--- eteam2 >  5.00
|   |   |   |   |   |   |   |   |--- eteam2 <= 7.00
|   |   |   |   |   |   |   |   |   |--- evenue <= 15.00
|   |   |   |   |   |   |   |   |   |   |--- evenue <= 5.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |   |   |--- evenue >  5.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |   |--- evenue >  15.00
|   |   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.45
|   |   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |   |   |--- opp_strength >  0.45
|   |   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |--- eteam2 >  7.00
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.52
|   |   |   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.52
|   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |--- eteam1 >  10.50
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.49
|   |   |   |   |   |   |   |   |--- eteam2 <= 4.00
|   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |--- eteam2 >  4.00
|   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.49
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.55
|   |   |   |   |   |   |   |   |   |--- evenue <= 7.00
|   |   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.51
|   |   |   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.51
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |--- evenue >  7.00
|   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |--- opp_strength >  0.55
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |--- eteam2 >  8.50
|   |   |   |   |   |--- toss_winner_strength <= 0.50
|   |   |   |   |   |   |--- opp_strength <= 0.51
|   |   |   |   |   |   |   |--- eteam1 <= 5.50
|   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |--- eteam1 >  5.50
```

```
|   |   |   |   |   |   |   |   |--- eteam1 <= 7.50
|   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |--- eteam1 >  7.50
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |--- opp_strength >  0.51
|   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |--- toss_winner_strength >  0.50
|   |   |   |   |   |   |--- eteam1 <= 4.00
|   |   |   |   |   |   |   |--- eteam1 <= 0.50
|   |   |   |   |   |   |   |   |--- class: Pune Warriors
|   |   |   |   |   |   |   |--- eteam1 >  0.50
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.28
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |--- opp_strength >  0.28
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.52
|   |   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.52
|   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |--- eteam1 >  4.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.48
|   |   |   |   |   |   |   |   |--- evenue <= 6.00
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |--- evenue >  6.00
|   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |--- opp_strength >  0.48
|   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.53
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.53
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |--- eteam1 >  12.50
|   |   |   |   |--- eteam1 <= 13.50
|   |   |   |   |   |--- toss_winner_strength <= 0.51
|   |   |   |   |   |   |--- eteam2 <= 9.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.50
|   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.48
|   |   |   |   |   |   |   |   |   |--- eteam2 <= 5.50
|   |   |   |   |   |   |   |   |   |   |--- class: Gujarat Lions
|   |   |   |   |   |   |   |   |   |--- eteam2 >  5.50
|   |   |   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.48
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.47
|   |   |   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.47
|   |   |   |   |   |   |   |   |   |   |--- eteam2 <= 3.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |   |   |--- eteam2 >  3.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |--- opp_strength >  0.50
```

```
|   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |--- eteam2 >  9.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.47
|   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.47
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.47
|   |   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |   |--- opp_strength >  0.47
|   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |--- toss_winner_strength >  0.51
|   |   |   |   |   |   |--- evenue <= 33.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.49
|   |   |   |   |   |   |   |   |--- evenue <= 24.50
|   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |--- evenue >  24.50
|   |   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |   |--- opp_strength >  0.49
|   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |--- evenue >  33.00
|   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |--- eteam1 >  13.50
|   |   |   |   |   |--- toss_winner_strength <= 0.47
|   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |--- toss_winner_strength >  0.47
|   |   |   |   |   |   |--- eteam2 <= 6.50
|   |   |   |   |   |   |   |--- eteam2 <= 5.00
|   |   |   |   |   |   |   |   |--- evenue <= 24.00
|   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |   |--- evenue >  24.00
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.51
|   |   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.51
|   |   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |--- eteam2 >  5.00
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.52
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.50
|   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.50
|   |   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |   |--- opp_strength >  0.52
|   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |--- eteam2 >  6.50
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.49
|   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.49
|   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |--- eteam2 >  10.50
|   |   |   |--- eteam2 <= 11.50
```

```
|   |   |   |   |--- eteam1 <= 11.50
|   |   |   |   |   |--- toss_winner_strength <= 0.52
|   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |--- toss_winner_strength >  0.52
|   |   |   |   |   |   |--- opp_strength <= 0.53
|   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |--- opp_strength >  0.53
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.53
|   |   |   |   |   |   |   |   |--- eteam1 <= 1.00
|   |   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |   |   |--- eteam1 >  1.00
|   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.53
|   |   |   |   |   |   |   |   |--- evenue <= 29.00
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |--- evenue >  29.00
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |--- eteam1 >  11.50
|   |   |   |   |   |--- opp_strength <= 0.49
|   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |--- opp_strength >  0.49
|   |   |   |   |   |   |--- eteam1 <= 13.50
|   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |--- eteam1 >  13.50
|   |   |   |   |   |   |   |--- opp_strength <= 0.53
|   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |--- opp_strength >  0.53
|   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |--- eteam2 >  11.50
|   |   |   |   |--- eteam2 <= 13.50
|   |   |   |   |   |--- eteam1 <= 9.00
|   |   |   |   |   |   |--- toss_winner_strength <= 0.52
|   |   |   |   |   |   |   |--- evenue <= 5.50
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.43
|   |   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |   |   |--- opp_strength >  0.43
|   |   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |   |--- evenue >  5.50
|   |   |   |   |   |   |   |   |--- evenue <= 16.00
|   |   |   |   |   |   |   |   |   |--- evenue <= 13.00
|   |   |   |   |   |   |   |   |   |   |--- evenue <= 10.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Royal Challengers
Bangalore
|   |   |   |   |   |   |   |   |   |   |--- evenue >  10.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |   |   |   |   |--- evenue >  13.00
|   |   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |   |--- evenue >  16.00
```

```
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.46
|   |   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.44
|   |   |   |   |   |   |   |   |   |   |   |--- class: Royal Challengers
Bangalore
|   |   |   |   |   |   |   |   |   |   |--- opp_strength >  0.44
|   |   |   |   |   |   |   |   |   |   |   |--- class: Kochi Tuskers Kerala
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.46
|   |   |   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |--- toss_winner_strength >  0.52
|   |   |   |   |   |   |   |--- evenue <= 22.00
|   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |--- evenue >  22.00
|   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |--- eteam1 >  9.00
|   |   |   |   |   |   |--- toss_winner_strength <= 0.51
|   |   |   |   |   |   |   |--- eteam1 <= 10.50
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.49
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.46
|   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.46
|   |   |   |   |   |   |   |   |   |   |--- eteam2 <= 12.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Rising Pune Supergiant
|   |   |   |   |   |   |   |   |   |   |--- eteam2 >  12.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Royal Challengers
Bangalore
|   |   |   |   |   |   |   |   |--- opp_strength >  0.49
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |--- eteam1 >  10.50
|   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |--- toss_winner_strength >  0.51
|   |   |   |   |   |   |   |--- opp_strength <= 0.49
|   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |--- opp_strength >  0.49
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.51
|   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |   |--- opp_strength >  0.51
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |--- eteam2 >  13.50
|   |   |   |   |   |--- eteam1 <= 9.00
|   |   |   |   |   |   |--- eteam1 <= 5.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.64
|   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |--- opp_strength >  0.64
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.69
|   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |--- opp_strength >  0.69
|   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |--- eteam1 >  5.00
```

```
|   |   |   |   |   |   |   |--- opp_strength <= 0.50
|   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |   |--- opp_strength >  0.50
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.66
|   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |   |--- opp_strength >  0.66
|   |   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |--- eteam1 >  9.00
|   |   |   |   |   |   |--- toss_winner_strength <= 0.48
|   |   |   |   |   |   |   |--- evenue <= 32.50
|   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |--- evenue >  32.50
|   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|   |   |   |   |   |   |--- toss_winner_strength >  0.48
|   |   |   |   |   |   |   |--- opp_strength <= 0.50
|   |   |   |   |   |   |   |   |--- evenue <= 23.00
|   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |   |--- evenue >  23.00
|   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.48
|   |   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |   |   |--- opp_strength >  0.48
|   |   |   |   |   |   |   |   |   |   |--- eteam1 <= 10.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |   |   |--- eteam1 >  10.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |--- opp_strength >  0.50
|   |   |   |   |   |   |   |   |--- eteam1 <= 12.00
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |--- eteam1 >  12.00
|   |   |   |   |   |   |   |   |   |--- class: Royal Challengers Bangalore
|--- toss_winner_strength >  0.54
|   |--- toss_winner_strength <= 0.59
|   |   |--- eteam2 <= 7.50
|   |   |   |--- eteam1 <= 7.50
|   |   |   |   |--- eteam2 <= 6.50
|   |   |   |   |   |--- eteam1 <= 6.50
|   |   |   |   |   |   |--- evenue <= 22.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.38
|   |   |   |   |   |   |   |   |--- eteam1 <= 1.00
|   |   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |   |   |--- eteam1 >  1.00
|   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |--- opp_strength >  0.38
|   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |--- evenue >  22.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.64
|   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |--- opp_strength >  0.64
```

```
|   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |--- eteam1 >  6.50
|   |   |   |   |   |   |--- eteam2 <= 1.00
|   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |--- eteam2 >  1.00
|   |   |   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |   |--- eteam2 >  6.50
|   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |--- eteam1 >  7.50
|   |   |   |   |--- eteam1 <= 8.50
|   |   |   |   |   |--- eteam2 <= 0.50
|   |   |   |   |   |   |--- opp_strength <= 0.57
|   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |--- opp_strength >  0.57
|   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |--- eteam2 >  0.50
|   |   |   |   |   |   |--- eteam2 <= 6.50
|   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |--- eteam2 >  6.50
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.57
|   |   |   |   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.57
|   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |--- eteam1 >  8.50
|   |   |   |   |   |--- opp_strength <= 0.51
|   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |--- opp_strength >  0.51
|   |   |   |   |   |   |--- eteam1 <= 12.50
|   |   |   |   |   |   |   |--- eteam2 <= 1.00
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.56
|   |   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |   |   |--- opp_strength >  0.56
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |--- eteam2 >  1.00
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.56
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |--- opp_strength >  0.56
|   |   |   |   |   |   |   |   |   |--- eteam2 <= 3.50
|   |   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |   |--- eteam2 >  3.50
|   |   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |--- eteam1 >  12.50
|   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |--- eteam2 >  7.50
|   |   |   |--- eteam2 <= 8.50
|   |   |   |   |--- toss_winner_strength <= 0.54
|   |   |   |   |   |--- opp_strength <= 0.50
|   |   |   |   |   |   |--- class: Mumbai Indians
```

```
|   |   |   |   |   |--- opp_strength >  0.50
|   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |--- toss_winner_strength >  0.54
|   |   |   |   |   |--- opp_strength <= 0.56
|   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |--- opp_strength >  0.56
|   |   |   |   |   |   |--- eteam1 <= 9.00
|   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |--- eteam1 >  9.00
|   |   |   |   |   |   |   |--- eteam1 <= 12.50
|   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |--- eteam1 >  12.50
|   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |--- eteam2 >  8.50
|   |   |   |   |--- toss_winner_strength <= 0.56
|   |   |   |   |   |--- eteam2 <= 11.50
|   |   |   |   |   |   |--- evenue <= 5.50
|   |   |   |   |   |   |   |--- evenue <= 2.00
|   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |--- evenue >  2.00
|   |   |   |   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |   |   |   |--- evenue >  5.50
|   |   |   |   |   |   |   |--- opp_strength <= 0.60
|   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |--- opp_strength >  0.60
|   |   |   |   |   |   |   |   |--- match_importance_code <= 0.50
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.55
|   |   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.55
|   |   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |--- match_importance_code >  0.50
|   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |--- eteam2 >  11.50
|   |   |   |   |   |   |--- evenue <= 32.50
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.56
|   |   |   |   |   |   |   |   |--- evenue <= 6.00
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |--- evenue >  6.00
|   |   |   |   |   |   |   |   |   |--- evenue <= 28.50
|   |   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |   |   |--- evenue >  28.50
|   |   |   |   |   |   |   |   |   |   |--- opp_strength <= 0.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |   |   |--- opp_strength >  0.50
|   |   |   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.56
|   |   |   |   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |   |   |   |--- evenue >  32.50
```

```
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.54
|   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.54
|   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |--- toss_winner_strength >  0.56
|   |   |   |   |   |--- eteam1 <= 4.00
|   |   |   |   |   |   |--- eteam1 <= 0.50
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.59
|   |   |   |   |   |   |   |   |--- eteam2 <= 13.50
|   |   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |   |   |--- eteam2 >  13.50
|   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.59
|   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |--- eteam1 >  0.50
|   |   |   |   |   |   |   |--- opp_strength <= 0.53
|   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.58
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.58
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |--- opp_strength >  0.53
|   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |--- eteam1 >  4.00
|   |   |   |   |   |   |--- eteam1 <= 9.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.56
|   |   |   |   |   |   |   |   |--- eteam1 <= 7.00
|   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |   |   |--- eteam1 >  7.00
|   |   |   |   |   |   |   |   |   |--- class: Mumbai Indians
|   |   |   |   |   |   |   |--- opp_strength >  0.56
|   |   |   |   |   |   |   |   |--- eteam2 <= 12.50
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |--- eteam2 >  12.50
|   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |   |   |--- eteam1 >  9.00
|   |   |   |   |   |   |   |--- eteam1 <= 12.00
|   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |--- eteam1 >  12.00
|   |   |   |   |   |   |   |   |--- opp_strength <= 0.45
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |   |   |--- opp_strength >  0.45
|   |   |   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |--- toss_winner_strength >  0.59
|   |   |--- toss_winner_strength <= 0.62
|   |   |   |--- toss_winner_strength <= 0.60
|   |   |   |   |--- eteam1 <= 1.00
|   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |--- eteam1 >  1.00
```

```
|   |   |   |   |   |--- opp_strength <= 0.46
|   |   |   |   |   |   |--- eteam2 <= 3.00
|   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |--- eteam2 >  3.00
|   |   |   |   |   |   |   |--- toss_winner_strength <= 0.60
|   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |--- toss_winner_strength >  0.60
|   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |--- opp_strength >  0.46
|   |   |   |   |   |   |--- evenue <= 10.50
|   |   |   |   |   |   |   |--- class: Gujarat Lions
|   |   |   |   |   |   |--- evenue >  10.50
|   |   |   |   |   |   |   |--- eteam1 <= 6.00
|   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |--- eteam1 >  6.00
|   |   |   |   |   |   |   |   |--- evenue <= 32.00
|   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |--- evenue >  32.00
|   |   |   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |--- toss_winner_strength >  0.60
|   |   |   |   |--- eteam2 <= 10.50
|   |   |   |   |   |--- evenue <= 13.50
|   |   |   |   |   |   |--- eteam2 <= 3.00
|   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |--- eteam2 >  3.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.54
|   |   |   |   |   |   |   |   |--- eteam1 <= 4.50
|   |   |   |   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |   |   |   |--- eteam1 >  4.50
|   |   |   |   |   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |   |   |   |   |--- opp_strength >  0.54
|   |   |   |   |   |   |   |   |--- evenue <= 8.00
|   |   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |   |   |--- evenue >  8.00
|   |   |   |   |   |   |   |   |   |--- class: Gujarat Lions
|   |   |   |   |   |--- evenue >  13.50
|   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |--- eteam2 >  10.50
|   |   |   |   |   |--- evenue <= 24.00
|   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |--- evenue >  24.00
|   |   |   |   |   |   |--- opp_strength <= 0.52
|   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |--- opp_strength >  0.52
|   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |--- toss_winner_strength >  0.62
|   |   |   |--- opp_strength <= 0.43
|   |   |   |   |--- toss_winner_strength <= 0.78
```

```
|   |   |   |   |   |--- opp_strength <= 0.09
|   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |--- opp_strength >  0.09
|   |   |   |   |   |   |--- evenue <= 28.00
|   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |   |--- evenue >  28.00
|   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |--- toss_winner_strength >  0.78
|   |   |   |   |   |--- eteam1 <= 11.00
|   |   |   |   |   |   |--- class: Delhi Capitals
|   |   |   |   |   |--- eteam1 >  11.00
|   |   |   |   |   |   |--- eteam1 <= 12.50
|   |   |   |   |   |   |   |--- class: Gujarat Lions
|   |   |   |   |   |   |--- eteam1 >  12.50
|   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |--- opp_strength >  0.43
|   |   |   |   |--- toss_winner_strength <= 0.65
|   |   |   |   |   |--- eteam1 <= 10.50
|   |   |   |   |   |   |--- toss_winner_strength <= 0.63
|   |   |   |   |   |   |   |--- evenue <= 22.50
|   |   |   |   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |   |   |   |   |--- evenue >  22.50
|   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |--- toss_winner_strength >  0.63
|   |   |   |   |   |   |   |--- class: Punjab Kings
|   |   |   |   |   |--- eteam1 >  10.50
|   |   |   |   |   |   |--- opp_strength <= 0.53
|   |   |   |   |   |   |   |--- evenue <= 22.50
|   |   |   |   |   |   |   |   |--- class: Gujarat Lions
|   |   |   |   |   |   |   |--- evenue >  22.50
|   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |--- opp_strength >  0.53
|   |   |   |   |   |   |   |--- class: Sunrisers Hyderabad
|   |   |   |   |--- toss_winner_strength >  0.65
|   |   |   |   |   |--- opp_strength <= 0.68
|   |   |   |   |   |   |--- eteam2 <= 8.00
|   |   |   |   |   |   |   |--- opp_strength <= 0.50
|   |   |   |   |   |   |   |   |--- eteam2 <= 5.50
|   |   |   |   |   |   |   |   |   |--- evenue <= 32.00
|   |   |   |   |   |   |   |   |   |   |--- class: Gujarat Titans
|   |   |   |   |   |   |   |   |   |--- evenue >  32.00
|   |   |   |   |   |   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |   |   |   |   |   |--- eteam2 >  5.50
|   |   |   |   |   |   |   |   |   |--- class: Lucknow Super Giants
|   |   |   |   |   |   |   |--- opp_strength >  0.50
|   |   |   |   |   |   |   |   |--- eteam1 <= 9.50
|   |   |   |   |   |   |   |   |   |--- class: Gujarat Titans
|   |   |   |   |   |   |   |   |--- eteam1 >  9.50
```

```
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.70
|   |   |   |   |   |   |   |   |   |   |--- class: Gujarat Titans
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.70
|   |   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |--- eteam2 >  8.00
|   |   |   |   |   |   |   |--- eteam2 <= 10.50
|   |   |   |   |   |   |   |   |--- eteam2 <= 9.50
|   |   |   |   |   |   |   |   |   |--- class: Pune Warriors
|   |   |   |   |   |   |   |   |--- eteam2 >  9.50
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength <= 0.87
|   |   |   |   |   |   |   |   |   |   |--- class: Gujarat Lions
|   |   |   |   |   |   |   |   |   |--- toss_winner_strength >  0.87
|   |   |   |   |   |   |   |   |   |   |--- class: Gujarat Titans
|   |   |   |   |   |   |   |--- eteam2 >  10.50
|   |   |   |   |   |   |   |   |--- evenue <= 37.00
|   |   |   |   |   |   |   |   |   |--- class: Rajasthan Royals
|   |   |   |   |   |   |   |   |--- evenue >  37.00
|   |   |   |   |   |   |   |   |   |--- class: Gujarat Titans
|   |   |   |   |   |--- opp_strength >  0.68
|   |   |   |   |   |   |--- eteam1 <= 1.50
|   |   |   |   |   |   |   |--- eteam2 <= 4.00
|   |   |   |   |   |   |   |   |--- class: Chennai Super Kings
|   |   |   |   |   |   |   |--- eteam2 >  4.00
|   |   |   |   |   |   |   |   |--- class: Kolkata Knight Riders
|   |   |   |   |   |   |--- eteam1 >  1.50
|   |   |   |   |   |   |   |--- class: Rising Pune Supergiant
```

10 Question 10: Advanced Match Prediction - Ensemble Methods (10 points)

10.0.1 Task:

Compare **Bagging** and **Boosting** for match outcome prediction:

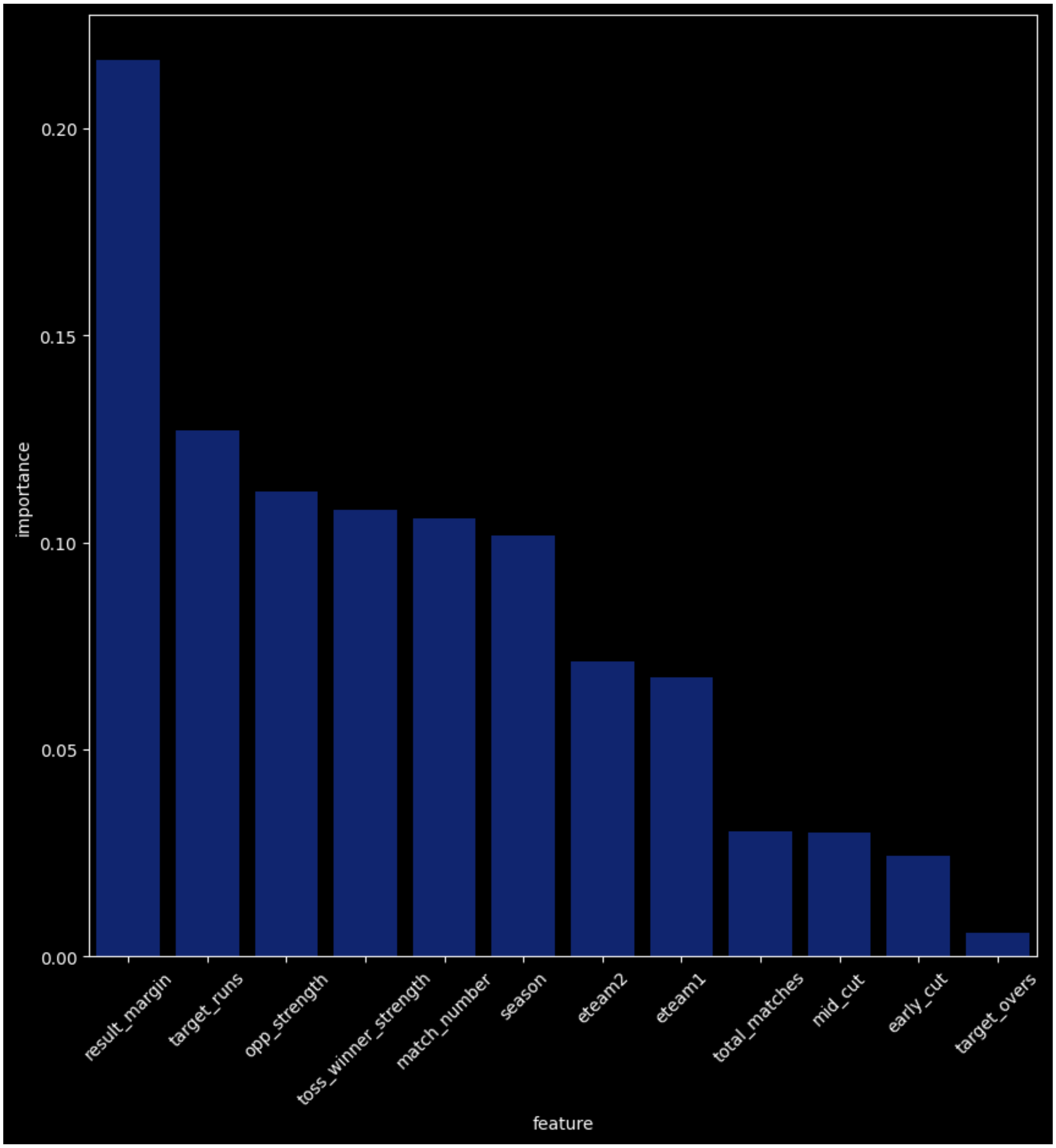
1. Implement Random Forest (Bagging):
   - Use all available features
   - Tune hyperparameters
   - Feature importance analysis
2. Implement XGBoost/AdaBoost (Boosting):
   - Compare with Random Forest
   - Analyze prediction confidence
3. Create an ensemble combining both approaches
4. Performance comparison:
   - Accuracy, Precision, Recall, F1-score

- ROC curves for all models

[55]:
```
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier,␣
↪VotingClassifier
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score,␣
↪f1_score, roc_auc_score, roc_curve
```

[56]:
```
xdf = pd.concat([df.select_dtypes(include = ['number' , 'float64' , 'int64']) ␣
↪, ddf.iloc[: , 4:]] , axis = 1)
xdf = xdf.drop(columns = [ 'id' , 'win'])
xtrain , xtest , ytrain , ytest = train_test_split(xdf , df['win'] , test_size=␣
↪0.2 , random_state= 42)
```

[57]:
```
rf = RandomForestClassifier(random_state= 42)
param_grid_rf = {
    'n_estimators': [100, 200],
    'max_depth': [5, 10, None],
    'min_samples_split': [2, 5],
}
grid_rf = GridSearchCV(rf, param_grid_rf, cv=3, scoring='accuracy', n_jobs=-1)
grid_rf.fit(xtrain, ytrain)
rf_best = grid_rf.best_estimator_

# Feature importance analysis
feature_importance = pd.DataFrame({
    'feature': xtrain.columns,
    'importance': rf_best.feature_importances_
}).sort_values(by='importance', ascending=False)
plt.figure(figsize = (10 , 10))
sns.barplot(data = feature_importance , x = 'feature' , y = 'importance')
plt.xticks(rotation = 45)
plt.show()
```



[58]:
```
#2. XGBoost:
xgb = XGBClassifier(use_label_encoder=False, eval_metric='logloss',␣
↪random_state=42)
xgb.fit(xtrain, ytrain)

adb = AdaBoostClassifier(random_state=42, n_estimators=100)
adb.fit(xtrain, ytrain)

# 3 Ensemble combining Bagging and Boosting
ensemble = VotingClassifier(estimators=[
```

```
    ('rf', rf_best),
    ('xgb', xgb),
    ('adb', adb)
], voting='soft')
ensemble.fit(xtrain, ytrain)
```

```
/opt/anaconda3/envs/tensorflow/lib/python3.11/site-
packages/xgboost/training.py:199: UserWarning: [16:15:48] WARNING:
/Users/runner/work/xgboost/xgboost/src/learner.cc:790:
Parameters: { "use_label_encoder" } are not used.

  bst.update(dtrain, iteration=i, fobj=obj)
/opt/anaconda3/envs/tensorflow/lib/python3.11/site-
packages/xgboost/training.py:199: UserWarning: [16:15:48] WARNING:
/Users/runner/work/xgboost/xgboost/src/learner.cc:790:
Parameters: { "use_label_encoder" } are not used.

  bst.update(dtrain, iteration=i, fobj=obj)
```

[58]:
```
VotingClassifier(estimators=[('rf', RandomForestClassifier(random_state=42)),
                             ('xgb',
                              XGBClassifier(base_score=None, booster=None,
                                            callbacks=None,
                                            colsample_bylevel=None,
                                            colsample_bynode=None,
                                            colsample_bytree=None, device=None,
                                            early_stopping_rounds=None,
                                            enable_categorical=False,
                                            eval_metric='logloss',
                                            feature_types=None,
                                            feature_weights=None, gamma=None,
                                            grow_…
                                            interaction_constraints=None,
                                            learning_rate=None, max_bin=None,
                                            max_cat_threshold=None,
                                            max_cat_to_onehot=None,
                                            max_delta_step=None, max_depth=None,
                                            max_leaves=None,
                                            min_child_weight=None, missing=nan,
                                            monotone_constraints=None,
                                            multi_strategy=None,
                                            n_estimators=None, n_jobs=None,
                                            num_parallel_tree=None, …)),
                             ('adb',
                              AdaBoostClassifier(n_estimators=100,
                                                 random_state=42))],
                 voting='soft')
```

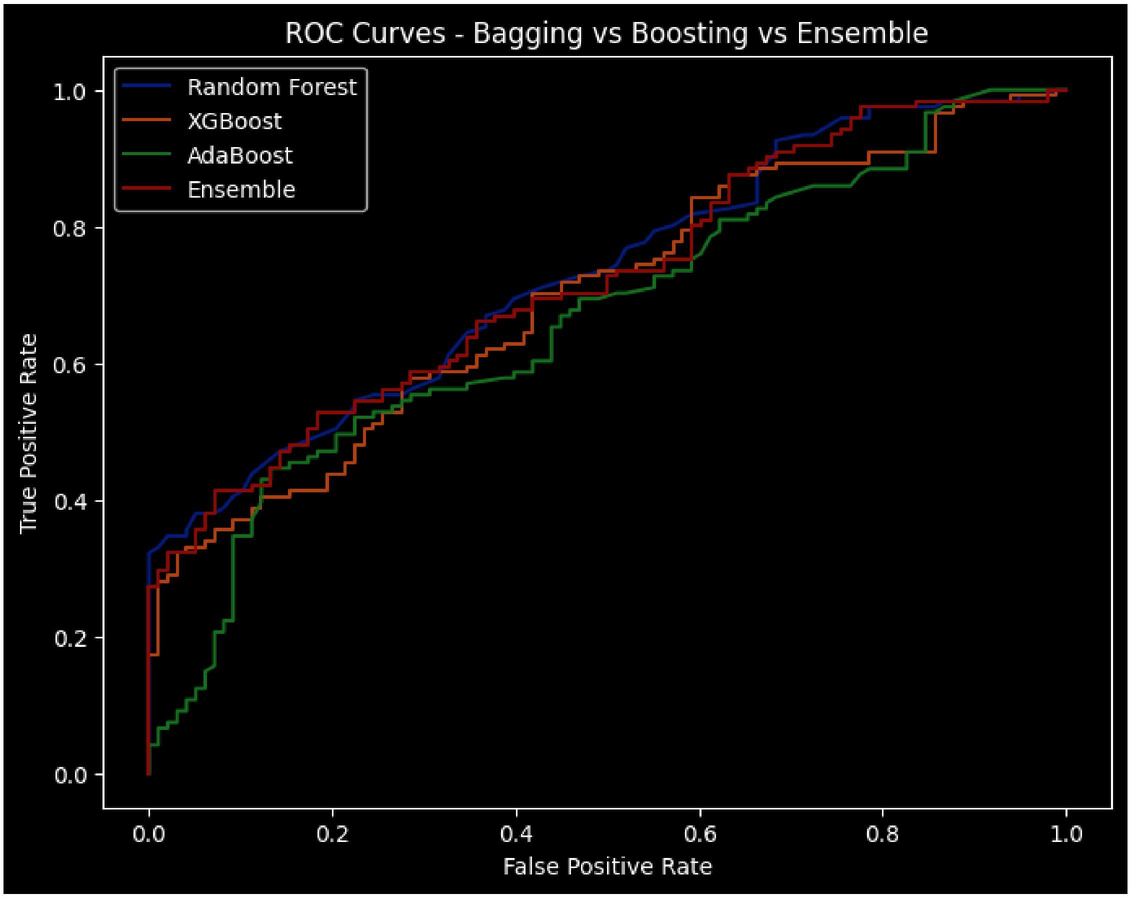
[69]:
```python
# import ace_tools as tools
# Predictions
models = {'Random Forest': rf_best, 'XGBoost': xgb, 'AdaBoost': adb, 'Ensemble':
 ↪ ensemble}
results = []

for name, model in models.items():
    y_pred = model.predict(xtest)
    y_prob = model.predict_proba(xtest)[:, 1] if hasattr(model,␣
 ↪'predict_proba') else np.zeros_like(y_pred)
    results.append({
        'Model': name,
        'Accuracy': accuracy_score(ytest, y_pred),
        'Precision': precision_score(ytest, y_pred, average='macro'),
        'Recall': recall_score(ytest, y_pred, average='macro'),
        'F1-Score': f1_score(ytest, y_pred, average='macro'),
        'ROC-AUC': roc_auc_score(ytest, y_prob, average='macro',␣
 ↪multi_class='ovr') if len(np.unique(y)) > 2 else roc_auc_score(ytest, y_prob)
    })

# Performance dataframe
perf_df = pd.DataFrame(results)

# 4 ROC Curves
plt.figure(figsize=(8,6))
for name, model in models.items():
    if hasattr(model, "predict_proba"):
        y_prob = model.predict_proba(xtest)[:, 1]
        # print(y_prob)
        print(ytest.shape , y_prob.shape)
        fpr, tpr, _ = roc_curve(ytest, y_prob)
        print(fpr , tpr)
        plt.plot(fpr, tpr, label=f'{name}')

# plt.plot([0, 1], [0, 1], 'k--')
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("ROC Curves - Bagging vs Boosting vs Ensemble")
plt.legend()
plt.show()

# Display feature importance and performance metrics
# tools.display_dataframe_to_user("Model Performance Comparison", perf_df)
# tools.display_dataframe_to_user("Random Forest Feature Importance",␣
 ↪feature_importance)
```

```
(219,) (219,)
```

```
[0.         0.         0.         0.         0.         0.
 0.         0.         0.         0.         0.         0.
 0.         0.         0.         0.         0.         0.01020408
 0.02040816 0.04081633 0.04081633 0.05102041 0.07142857 0.08163265
 0.09183673 0.10204082 0.1122449  0.14285714 0.20408163 0.2244898
 0.2244898  0.24489796 0.26530612 0.2755102  0.28571429 0.31632653
 0.32653061 0.34693878 0.36734694 0.36734694 0.3877551  0.39795918
 0.42857143 0.46938776 0.5        0.51020408 0.52040816 0.54081633
 0.55102041 0.57142857 0.59183673 0.63265306 0.66326531 0.66326531
 0.66326531 0.68367347 0.68367347 0.71428571 0.7244898  0.74489796
 0.75510204 0.78571429 0.78571429 0.82653061 0.85714286 0.86734694
 0.92857143 0.94897959 0.94897959 0.96938776 0.97959184 0.98979592
 1.        ] [0.         0.01652893 0.04958678 0.05785124 0.09090909 0.10743802
 0.11570248 0.14049587 0.15702479 0.18181818 0.19834711 0.21487603
 0.2231405  0.23966942 0.26446281 0.27272727 0.32231405 0.33057851
 0.34710744 0.34710744 0.3553719  0.38016529 0.38016529 0.38842975
 0.40495868 0.41322314 0.43801653 0.47107438 0.50413223 0.53719008
 0.54545455 0.55371901 0.55371901 0.55371901 0.56198347 0.5785124
 0.61157025 0.6446281  0.65289256 0.66942149 0.67768595 0.69421488
 0.7107438  0.72727273 0.73553719 0.74380165 0.76859504 0.7768595
 0.79338843 0.80165289 0.81818182 0.82644628 0.83471074 0.85950413
 0.87603306 0.90909091 0.92561983 0.9338843  0.9338843  0.95041322
 0.95867769 0.95867769 0.97520661 0.97520661 0.97520661 0.98347107
 0.98347107 0.98347107 0.99173554 0.99173554 0.99173554 1.
 1.        ]
(219,) (219,)
[0.         0.         0.         0.01020408 0.01020408 0.02040816
 0.02040816 0.03061224 0.03061224 0.04081633 0.04081633 0.06122449
 0.06122449 0.07142857 0.07142857 0.09183673 0.09183673 0.1122449
 0.1122449  0.12244898 0.12244898 0.15306122 0.15306122 0.19387755
 0.19387755 0.21428571 0.21428571 0.2244898  0.2244898  0.23469388
 0.23469388 0.24489796 0.24489796 0.25510204 0.25510204 0.2755102
 0.2755102  0.28571429 0.28571429 0.30612245 0.30612245 0.34693878
 0.34693878 0.35714286 0.35714286 0.36734694 0.36734694 0.3877551
 0.3877551  0.40816327 0.40816327 0.41836735 0.41836735 0.44897959
 0.44897959 0.46938776 0.46938776 0.48979592 0.48979592 0.53061224
 0.53061224 0.55102041 0.55102041 0.56122449 0.56122449 0.57142857
 0.57142857 0.58163265 0.58163265 0.59183673 0.59183673 0.62244898
 0.62244898 0.63265306 0.63265306 0.66326531 0.66326531 0.68367347
 0.68367347 0.78571429 0.78571429 0.85714286 0.85714286 0.87755102
 0.87755102 0.8877551  0.8877551  0.93877551 0.93877551 0.98979592
 0.98979592 1.        ] [0.         0.00826446 0.17355372 0.17355372 0.28099174
0.28099174
 0.2892562  0.2892562  0.32231405 0.32231405 0.33057851 0.33057851
 0.33884298 0.33884298 0.3553719  0.3553719  0.37190083 0.37190083
 0.38842975 0.38842975 0.40495868 0.40495868 0.41322314 0.41322314
 0.43801653 0.43801653 0.45454545 0.45454545 0.47933884 0.47933884
 0.50413223 0.50413223 0.51239669 0.51239669 0.52892562 0.52892562
```

```
 0.57024793 0.57024793 0.5785124  0.5785124  0.58677686 0.58677686
 0.59504132 0.59504132 0.61157025 0.61157025 0.61983471 0.61983471
 0.62809917 0.62809917 0.6446281  0.6446281  0.70247934 0.70247934
 0.71900826 0.71900826 0.72727273 0.72727273 0.73553719 0.73553719
 0.74380165 0.74380165 0.75206612 0.75206612 0.76033058 0.76033058
 0.7768595  0.7768595  0.79338843 0.79338843 0.84297521 0.84297521
 0.85950413 0.85950413 0.87603306 0.87603306 0.88429752 0.88429752
 0.89256198 0.89256198 0.90909091 0.90909091 0.96694215 0.96694215
 0.97520661 0.97520661 0.98347107 0.98347107 0.99173554 0.99173554
 1.         1.        ]
(219,) (219,)
[0.         0.         0.         0.01020408 0.01020408 0.02040816
 0.02040816 0.03061224 0.03061224 0.04081633 0.04081633 0.05102041
 0.05102041 0.06122449 0.06122449 0.07142857 0.07142857 0.07142857
 0.08163265 0.08163265 0.09183673 0.09183673 0.09183673 0.09183673
 0.09183673 0.1122449  0.1122449  0.12244898 0.12244898 0.12244898
 0.13265306 0.13265306 0.15306122 0.15306122 0.17346939 0.17346939
 0.18367347 0.18367347 0.20408163 0.20408163 0.2244898  0.2244898
 0.24489796 0.24489796 0.26530612 0.26530612 0.2755102  0.2755102
 0.28571429 0.28571429 0.30612245 0.30612245 0.34693878 0.34693878
 0.3877551  0.39795918 0.39795918 0.41836735 0.41836735 0.43877551
 0.43877551 0.44897959 0.44897959 0.45918367 0.45918367 0.46938776
 0.46938776 0.48979592 0.51020408 0.52040816 0.55102041 0.55102041
 0.57142857 0.57142857 0.59183673 0.59183673 0.60204082 0.6122449
 0.62244898 0.62244898 0.65306122 0.65306122 0.66326531 0.66326531
 0.67346939 0.67346939 0.68367347 0.7244898  0.76530612 0.7755102
 0.78571429 0.79591837 0.81632653 0.82653061 0.82653061 0.84693878
 0.84693878 0.84693878 0.84693878 0.85714286 0.86734694 0.87755102
 0.87755102 0.91836735 0.92857143 0.95918367 0.97959184 1.        ] [0.
0.00826446 0.04132231 0.04132231 0.0661157  0.0661157
 0.07438017 0.07438017 0.09090909 0.09090909 0.10743802 0.10743802
 0.12396694 0.12396694 0.14876033 0.15702479 0.19008264 0.20661157
 0.20661157 0.2231405  0.2231405  0.26446281 0.32231405 0.33884298
 0.34710744 0.34710744 0.37190083 0.39669421 0.41322314 0.42975207
 0.42975207 0.44628099 0.44628099 0.45454545 0.45454545 0.46280992
 0.46280992 0.47107438 0.47107438 0.49586777 0.49586777 0.52066116
 0.52066116 0.52892562 0.52892562 0.53719008 0.53719008 0.54545455
 0.54545455 0.55371901 0.55371901 0.56198347 0.56198347 0.57024793
 0.5785124  0.5785124  0.58677686 0.58677686 0.60330579 0.60330579
 0.65289256 0.65289256 0.66942149 0.66942149 0.67768595 0.67768595
 0.69421488 0.69421488 0.70247934 0.70247934 0.7107438  0.72727273
 0.72727273 0.73553719 0.73553719 0.75206612 0.76033058 0.78512397
 0.79338843 0.80991736 0.80991736 0.81818182 0.81818182 0.82644628
 0.82644628 0.83471074 0.84297521 0.85950413 0.85950413 0.87603306
 0.88429752 0.88429752 0.88429752 0.88429752 0.90909091 0.90909091
 0.92561983 0.95041322 0.96694215 0.96694215 0.97520661 0.97520661
 0.98347107 1.         1.         1.         1.         1.        ]
(219,) (219,)
```

```
[0.         0.         0.         0.01020408 0.01020408 0.02040816
 0.02040816 0.05102041 0.05102041 0.06122449 0.06122449 0.07142857
 0.07142857 0.1122449  0.1122449  0.13265306 0.13265306 0.14285714
 0.14285714 0.15306122 0.15306122 0.17346939 0.17346939 0.18367347
 0.18367347 0.2244898  0.2244898  0.25510204 0.25510204 0.2755102
 0.2755102  0.28571429 0.28571429 0.31632653 0.31632653 0.32653061
 0.32653061 0.33673469 0.33673469 0.34693878 0.34693878 0.35714286
 0.35714286 0.37755102 0.37755102 0.39795918 0.39795918 0.41836735
 0.41836735 0.44897959 0.44897959 0.5        0.5        0.51020408
 0.51020408 0.56122449 0.56122449 0.59183673 0.59183673 0.60204082
 0.60204082 0.6122449  0.6122449  0.63265306 0.63265306 0.65306122
 0.65306122 0.66326531 0.66326531 0.67346939 0.67346939 0.68367347
 0.68367347 0.70408163 0.70408163 0.74489796 0.74489796 0.75510204
 0.75510204 0.76530612 0.76530612 0.7755102  0.7755102  0.83673469
 0.83673469 0.97959184 0.97959184 1.        ] [0.         0.00826446 0.27272727
0.27272727 0.29752066 0.29752066
 0.32231405 0.32231405 0.3553719  0.3553719  0.38016529 0.38016529
 0.41322314 0.41322314 0.4214876  0.4214876  0.44628099 0.44628099
 0.47107438 0.47107438 0.47933884 0.47933884 0.50413223 0.50413223
 0.52892562 0.52892562 0.54545455 0.54545455 0.56198347 0.56198347
 0.57024793 0.57024793 0.58677686 0.58677686 0.59504132 0.59504132
 0.60330579 0.60330579 0.61157025 0.61157025 0.63636364 0.63636364
 0.66115702 0.66115702 0.66942149 0.66942149 0.67768595 0.67768595
 0.69421488 0.69421488 0.70247934 0.70247934 0.72727273 0.72727273
 0.73553719 0.73553719 0.75206612 0.75206612 0.80165289 0.80165289
 0.80991736 0.80991736 0.83471074 0.83471074 0.87603306 0.87603306
 0.88429752 0.88429752 0.89256198 0.89256198 0.90082645 0.90082645
 0.90909091 0.90909091 0.91735537 0.91735537 0.9338843  0.9338843
 0.94214876 0.94214876 0.95867769 0.95867769 0.97520661 0.97520661
 0.98347107 0.98347107 1.         1.        ]
```



[61]:
```
perf_df
```

[61]:

	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
0	Random Forest	0.630137	0.640383	0.639104	0.630014	0.729718
1	XGBoost	0.630137	0.640383	0.639104	0.630014	0.698516
2	AdaBoost	0.625571	0.649356	0.640791	0.623301	0.665078
3	Ensemble	0.634703	0.642555	0.642267	0.634696	0.721707

[ ]: