IPL Data Analysis

Introduction

The Indian Premier League, one of the most fast-moving and followed leagues for the game of cricket, was laced with high-energy matches and high-class performances since it finally came into the limelight. This project would delve deep into statistical aspects regarding top performers of IPL, the performance trends within each match and teams participating across the series, for this analysis.

We used Python and the Pandas library along with other visualization libraries like Matplotlib and Seaborn, analyzing an IPL dataset on the following terms:

- Who the top players are?
- Which teams have had the best performances all these years?
- What trends exist in the match outcomes?

```
#loading the required libraries

import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
```

Overview of Dataset

The dataset used for this analysis is very rich in IPL matches information, including:

Features: Match ID, season, city, winner, player of the match, team statistics, and many more.

```
from google.colab import drive

drive.mount('/content/drive')

# Path to the file
file_path = '/content/drive/My Drive/Project/IPL/matches.csv'

# Load the CSV into a DataFrame
ipl = pd.read_csv(file_path)
print(ipl.head())
```

```
Mounted at /content/drive
   id  season       city        date                       team1  \
0   1    2017  Hyderabad  2017-04-05          Sunrisers Hyderabad
1   2    2017       Pune  2017-04-06               Mumbai Indians
2   3    2017     Rajkot  2017-04-07                Gujarat Lions
3   4    2017     Indore  2017-04-08       Rising Pune Supergiant
4   5    2017  Bangalore  2017-04-08  Royal Challengers Bangalore

                        team2                 toss_winner toss_decision  \
0  Royal Challengers Bangalore  Royal Challengers Bangalore         field
1       Rising Pune Supergiant       Rising Pune Supergiant         field
2        Kolkata Knight Riders        Kolkata Knight Riders         field
3              Kings XI Punjab              Kings XI Punjab         field
4             Delhi Daredevils  Royal Challengers Bangalore           bat

   result  dl_applied                       winner  win_by_runs  \
0  normal           0          Sunrisers Hyderabad           35
1  normal           0       Rising Pune Supergiant            0
2  normal           0        Kolkata Knight Riders            0
3  normal           0              Kings XI Punjab            0
4  normal           0  Royal Challengers Bangalore           15

   win_by_wickets player_of_match                                      venue  \
0               0    Yuvraj Singh  Rajiv Gandhi International Stadium, Uppal
1               7       SPD Smith    Maharashtra Cricket Association Stadium
2              10         CA Lynn     Saurashtra Cricket Association Stadium
3               6      GJ Maxwell                      Holkar Cricket Stadium
4               0       KM Jadhav                       M Chinnaswamy Stadium

            umpire1        umpire2 umpire3
0     AY Dandekar       NJ Llong     NaN
1  A Nand Kishore         S Ravi     NaN
2     Nitin Menon      CK Nandan     NaN
3    AK Chaudhary  C Shamshuddin     NaN
4             NaN            NaN     NaN
```

Size: (756,18)

This dataset acts as a basis for finding important patterns and generating actionable insights

```
#Lookin at the number of rows and columns in the dataset
ipl.shape
```

```
(756, 18)
```

Analysis Highlights

1.Player Performers overview

The big players this IPL season are under review from different areas that highlight their performances to become big through things like numbers of "Player of the Match" awards, consistencies in key statistics, and an all-round contribution to teams' success. This information shows what it needs in that category to highlight who brings about the biggest difference on the cricket pitch with a view to measuring true brilliance in their attempt to inspire the outcome of any match.

Most "Player of the Match" Awards

One of the key metrics for individual performance is the "Player of the Match" award.

```
#Getting the frequency of most man of the match awards
ipl['player_of_match'].value_counts()
```

player_of_match	count
CH Gayle	21
AB de Villiers	20
RG Sharma	17
MS Dhoni	17
DA Warner	17
...	...
PD Collingwood	1
NV Ojha	1
AC Voges	1
J Theron	1
S Hetmyer	1

226 rows × 1 columns

dtype: int64

Here's a look at the top 10 players who have earned the most awards

```
#Getting the top 10 players with most man of the match awards
ipl['player_of_match'].value_counts()[0:10]
```

player_of_match	count
CH Gayle	21
AB de Villiers	20
RG Sharma	17
MS Dhoni	17
DA Warner	17
YK Pathan	16
SR Watson	15
SK Raina	14
G Gambhir	13
MEK Hussey	12

dtype: int64

Here's a look at the top 5 players who have earned the most awards:

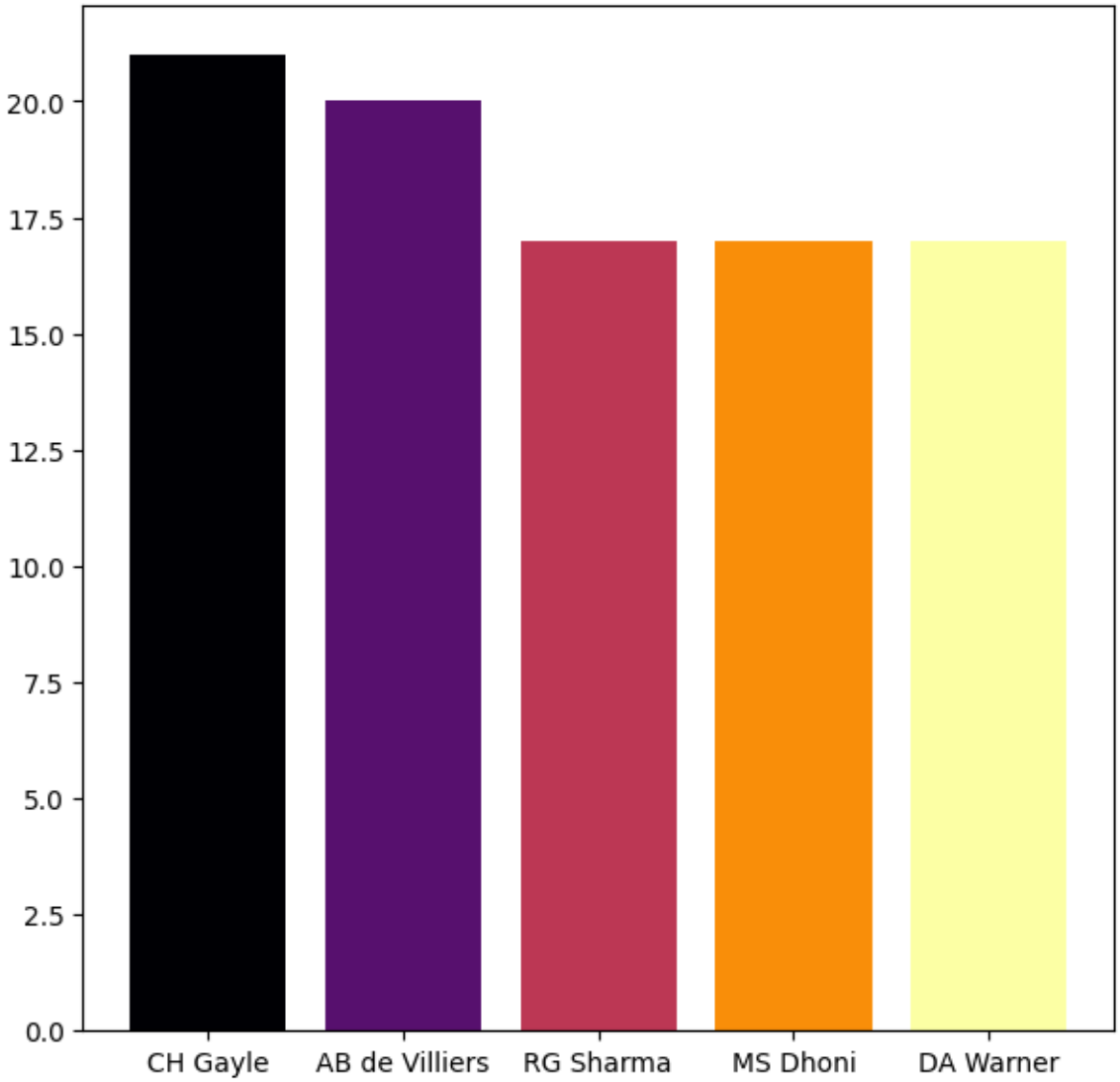
```
#Getting the top 5 players with most man of the match awards
ipl['player_of_match'].value_counts()[0:5]
```

player_of_match	count
CH Gayle	21
AB de Villiers	20
RG Sharma	17
MS Dhoni	17
DA Warner	17

dtype: int64

```
#making a bar-plot for the top 5 players with most man of the match awards
from matplotlib import cm

plt.figure(figsize=(7,7))
colors = cm.get_cmap('inferno', 5)
plt.bar(list(ipl['player_of_match'].value_counts()[0:5].keys()),list(ipl['
player_of_match'].value_counts()[0:5]), color = colors(range(5)))
plt.show()
```



2.Team Performance Overview

Overview of Team Performance reflects how teams have performed in different match scenarios, such as winning rates when batting first versus batting second. By looking at these statistics, we will understand the strengths and weaknesses of each team in various conditions, which can give us more insight into how strategy and match situation has influenced their success in the tournament.

```
#Getting the frequency of result column
ipl['result'].value_counts()
```

result	count
normal	743
tie	9
no result	4

dtype: int64

```
#Finding out the number of toss wins w.r.t each team
ipl['toss_winner'].value_counts()
```

toss_winner	count
Mumbai Indians	98
Kolkata Knight Riders	92
Chennai Super Kings	89
Royal Challengers Bangalore	81
Kings XI Punjab	81
Delhi Daredevils	80
Rajasthan Royals	80
Sunrisers Hyderabad	46
Deccan Chargers	43
Pune Warriors	20
Gujarat Lions	15
Delhi Capitals	10
Kochi Tuskers Kerala	8
Rising Pune Supergiants	7
Rising Pune Supergiant	6

dtype: int64

```
#Finding out how many times a team has won the match after winning the toss
import numpy as np
np.sum(ipl['toss_winner']==ipl['winner'])
```

```
393
```

```
325/636
```

```
0.5110062893081762
```

Batting First Analysis

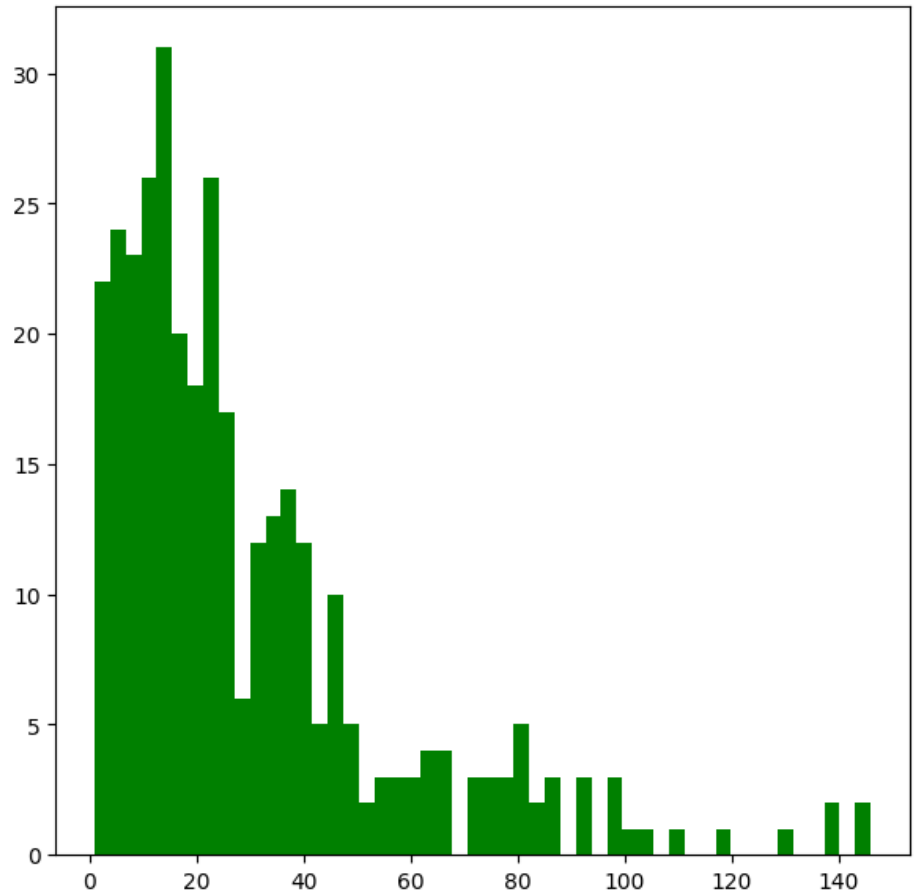
This section showcases teams that bat first in the IPL matches. We look at some of the performance metrics such as runs scored, wickets lost, and outcomes of matches to understand how teams perform when setting targets. By looking at past data, we try to find patterns that indicate the strengths and weaknesses of teams batting first.

```
#Extracting the records where a team won batting first
batting_first=ipl[ipl['win_by_runs']!=0]
```

```
#Looking at the head
batting_first.head()
```

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal	AY Dandekar
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0	KM Jadhav	M Chinnaswamy Stadium	NaN
8	9	2017	Pune	2017-04-11	Delhi Daredevils	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Delhi Daredevils	97	0	SV Samson	Maharashtra Cricket Association Stadium	AY Dandekar
13	14	2017	Kolkata	2017-04-15	Kolkata Knight Riders	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Kolkata Knight Riders	17	0	RV Uthappa	Eden Gardens	AY Dandekar
14	15	2017	Delhi	2017-04-15	Delhi Daredevils	Kings XI Punjab	Delhi Daredevils	bat	normal	0	Delhi Daredevils	51	0	CJ Anderson	Feroz Shah Kotla	YC Barde

```
#Making a histogram
plt.figure(figsize=(7,7))
plt.hist(batting_first['win_by_runs'],bins =50, color= 'green')
plt.show()
```

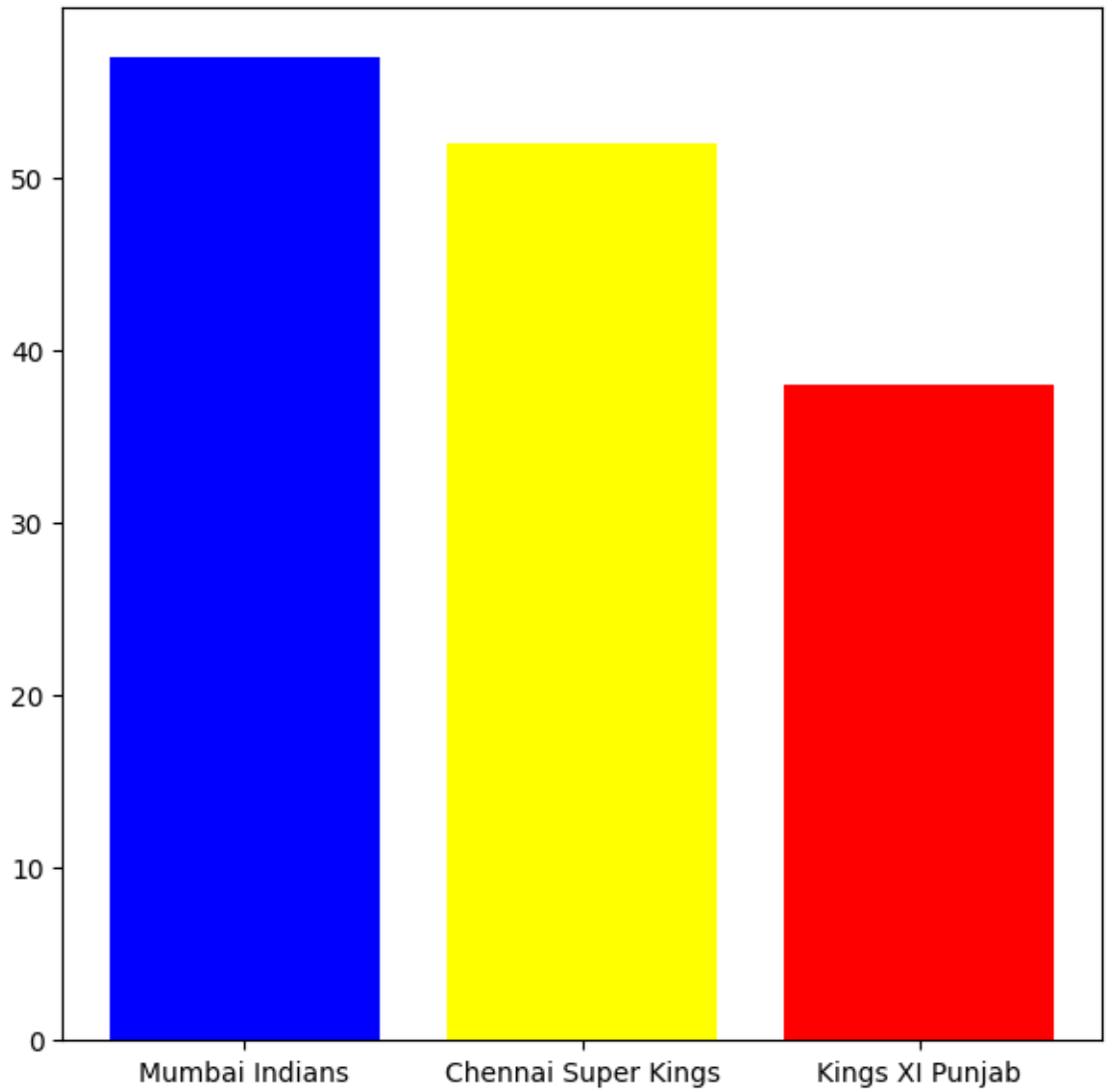



```
#Finding out the number of wins w.r.t each team after batting first
batting_first['winner'].value_counts()
```

winner	count
Mumbai Indians	57
Chennai Super Kings	52
Kings XI Punjab	38
Kolkata Knight Riders	36
Royal Challengers Bangalore	35
Sunrisers Hyderabad	30
Rajasthan Royals	27
Delhi Daredevils	25
Deccan Chargers	18
Pune Warriors	6
Rising Pune Supergiant	5
Delhi Capitals	3
Kochi Tuskers Kerala	2
Rising Pune Supergiants	2
Gujarat Lions	1

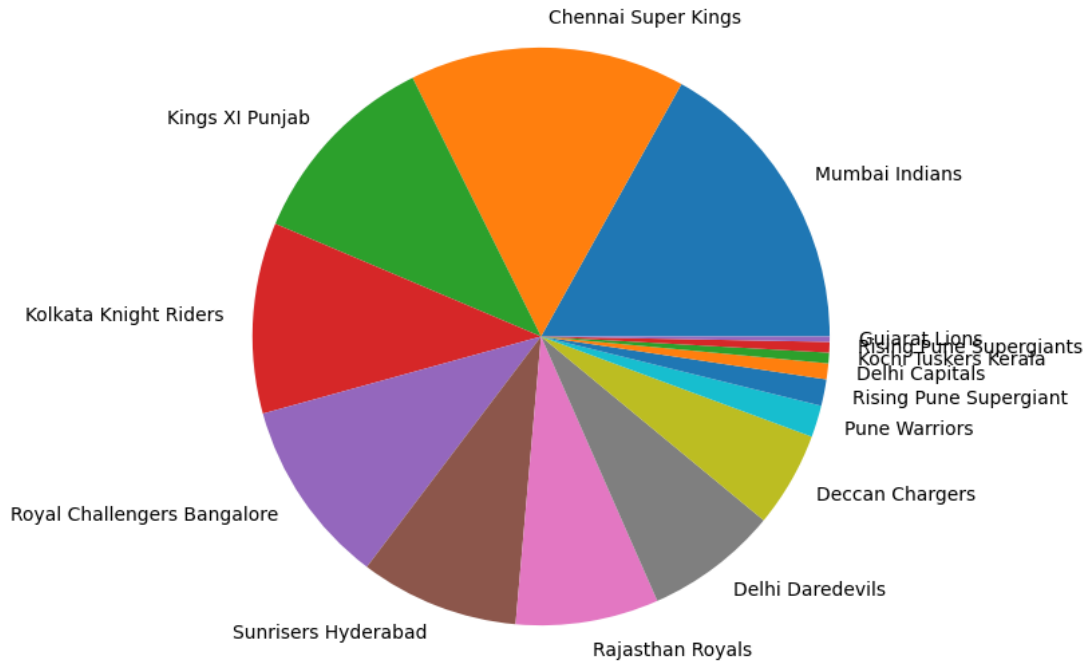
dtype: int64

```
#Making a bar-plot for top 3 teams with most wins after batting first
plt.figure(figsize=(7,7))
plt.bar(list(batting_first['winner'].value_counts()[0:3].keys()),list(batt
ing_first['winner'].value_counts()[0:3]),color=["blue","yellow","red"])
plt.show()
```




```
      #Making a pie chart

plt.figure(figsize=(7,7))
plt.pie(list(batting_first['winner'].value_counts()),labels=list(batting_f
irst['winner'].value_counts().keys()))
plt.show()
```



Batting Second Analysis

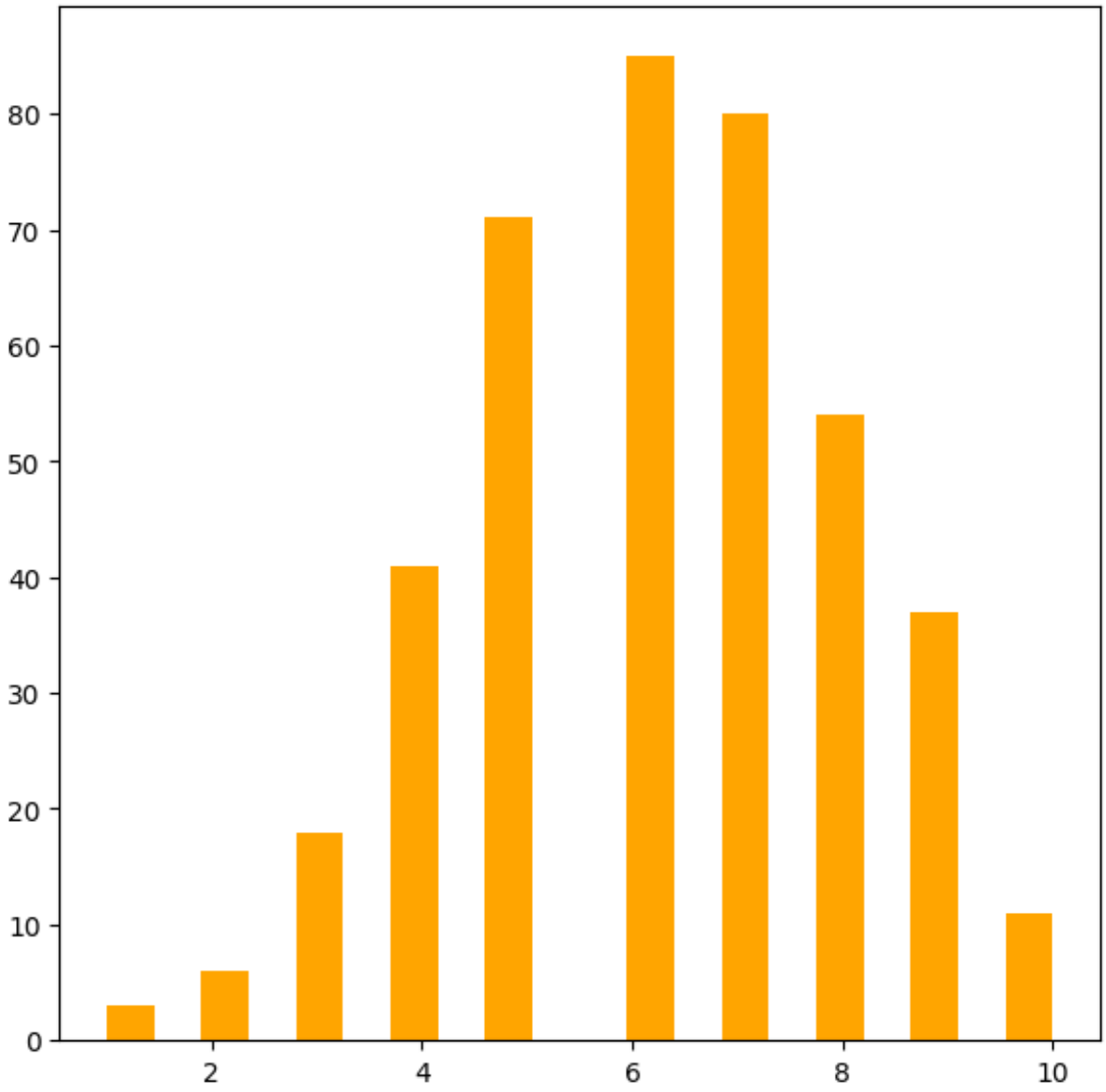
We now move on to the teams that bat second in IPL matches. We then go on to discuss how chasing teams respond to the target in terms of run rates, strategies to win matches, and the effect of chasing under different conditions. This will help us understand the psychological and strategic dynamics which come into play when teams chase a target

```
#extracting those records where a team has won after batting second
batting_second=ipl[ipl['win_by_wickets']!=0]
```

```
#looking at the head
batting_second.head()
```

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	0	7	SPD Smith	Maharashtra Cricket Association Stadium	A Nand Kishore	
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	10	CA Lynn	Saurashtra Cricket Association Stadium	Nitin Menon	C
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	6	GJ Maxwell	Holkar Cricket Stadium	AK Chaudhary	Sha
5	6	2017	Hyderabad	2017-04-09	Gujarat Lions	Sunrisers Hyderabad	Sunrisers Hyderabad	field	normal	0	Sunrisers Hyderabad	0	9	Rashid Khan	Rajiv Gandhi International Stadium, Uppal	A Deshmukh	
6	7	2017	Mumbai	2017-04-09	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field	normal	0	Mumbai Indians	0	4	N Rana	Wankhede Stadium	Nitin Menon	C

```
#Making a histogram for frequency of wins w.r.t number of wickets
plt.figure(figsize=(7,7))
plt.hist(batting_second['win_by_wickets'],bins=20,color='orange')
plt.show()
```

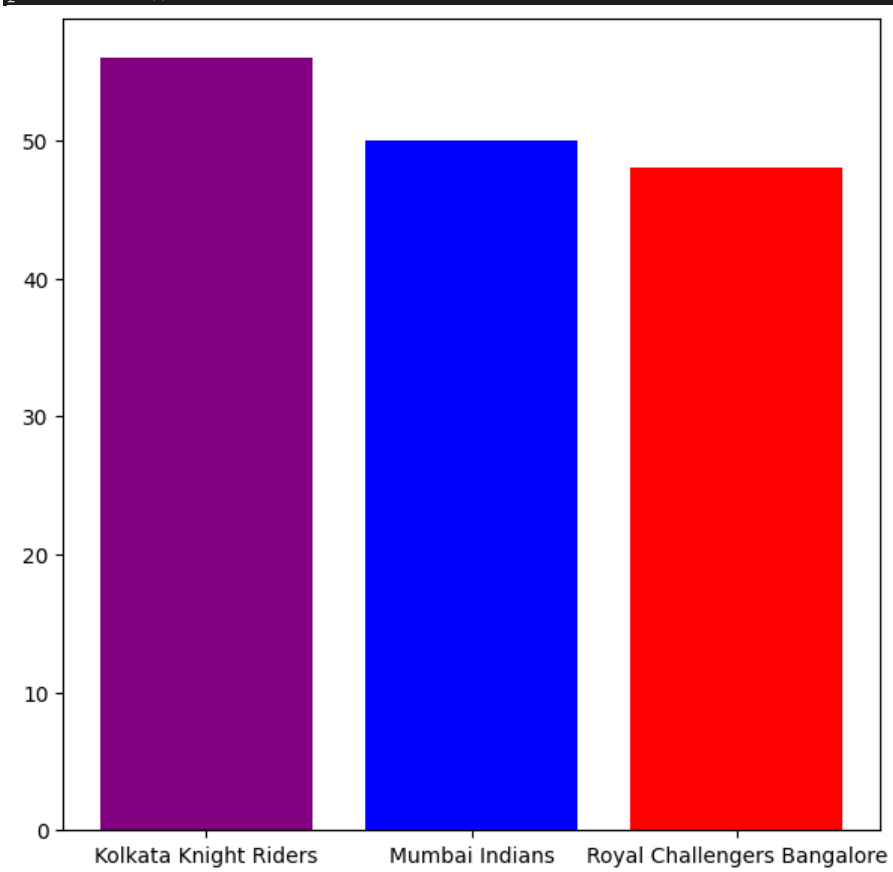



```
#Finding out the frequency of number of wins w.r.t each time after batting
second
batting_second['winner'].value_counts()
```

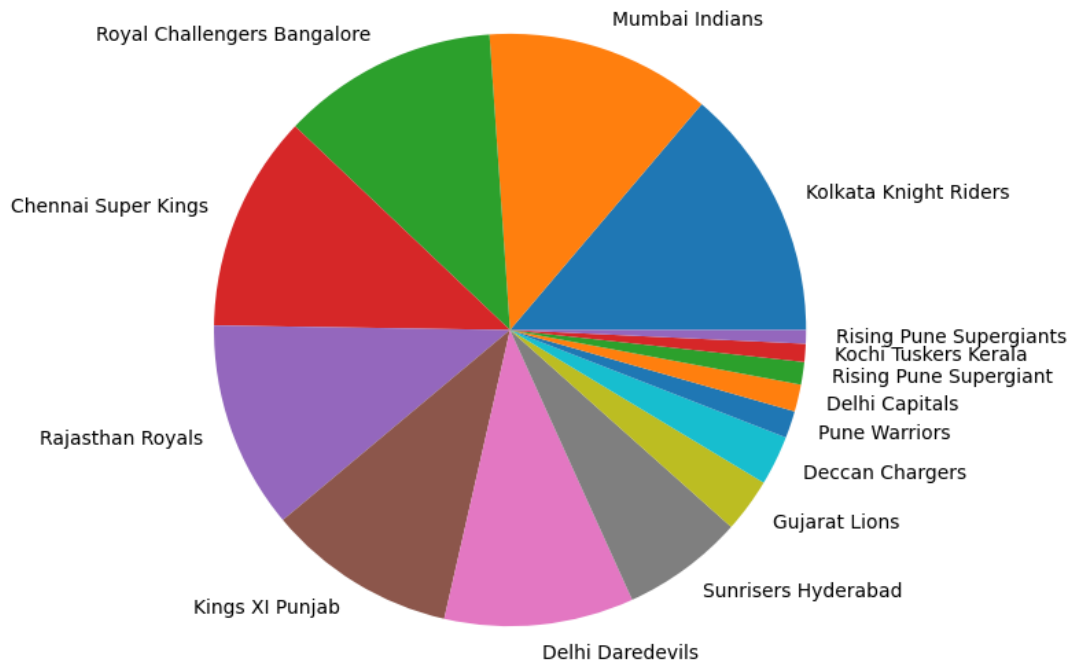
winner	count
Kolkata Knight Riders	56
Mumbai Indians	50
Royal Challengers Bangalore	48
Chennai Super Kings	48
Rajasthan Royals	46
Kings XI Punjab	42
Delhi Daredevils	42
Sunrisers Hyderabad	27
Gujarat Lions	12
Deccan Chargers	11
Pune Warriors	6
Delhi Capitals	6
Rising Pune Supergiant	5
Kochi Tuskers Kerala	4
Rising Pune Supergiants	3

dtype: int64

```
#Making a bar plot for top-3 teams with most wins after batting second
plt.figure(figsize=(7,7))
plt.bar(list(batting_second['winner'].value_counts()[0:3].keys()),list(bat
ting_second['winner'].value_counts()[0:3]),color=["purple","blue","red"])
plt.show()
```



```
#Making a pie chart for distribution of most wins after batting second
plt.figure(figsize=(7,7))
plt.pie(list(batting_second['winner'].value_counts()),labels=list(batting_
second['winner'].value_counts().keys()))
plt.show()
```



3.Match Trends

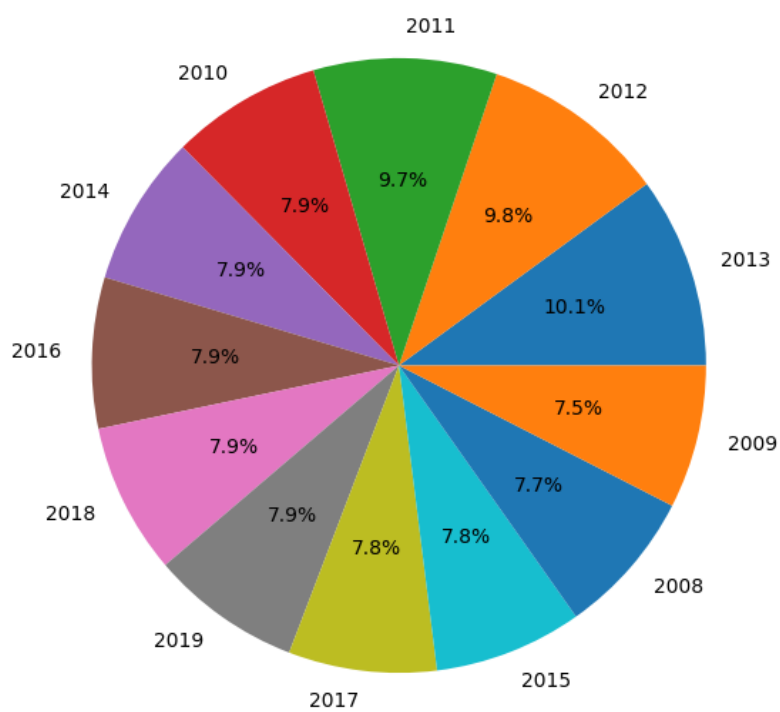
We observe the major trends of the IPL by considering the number of matches played during each season and across various cities. Examining the frequency of the games held in different locations and how they have evolved provides insight into the geographical dynamics of the tournament and how those might impact team performances and fan engagement. This helps outline patterns that may define the future seasons and give further context to the expanding footprint of the IPL.

```
#Looking at the number of matches played each season
ipl['season'].value_counts()
```

season	count
2013	76
2012	74
2011	73
2010	60
2014	60
2016	60
2018	60
2019	60
2017	59
2015	59
2008	58
2009	57

dtype: int64

```
#Making a pie chart for distribution the number of matches played each season
plt.figure(figsize=(7,7))
plt.pie(list(ipl['season'].value_counts()),labels=list(ipl['season'].value_counts().keys()),autopct='%0.1f%%')
plt.show()
```

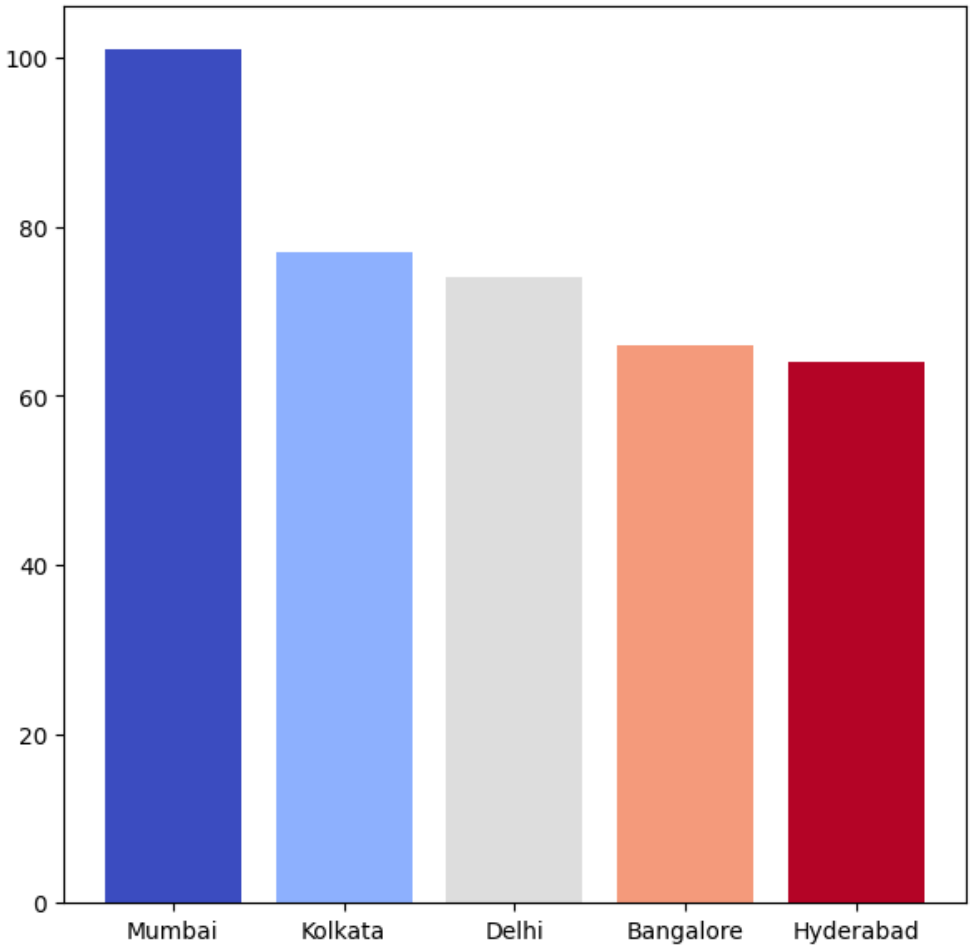


```
#Looking at the number of matches played in each city
ipl['city'].value_counts()
```

city	count
Mumbai	101
Kolkata	77
Delhi	74
Bangalore	66
Hyderabad	64
Chennai	57
Jaipur	47
Chandigarh	46
Pune	38
Durban	15
Bengaluru	14
Visakhapatnam	13
Centurion	12
Ahmedabad	12
Rajkot	10
Mohali	10
Indore	9
Dharamsala	9
Johannesburg	8

```
#making a bar-plot for the top 5 cities in the number of matches played
from matplotlib import cm

plt.figure(figsize=(7,7))
colors = cm.get_cmap('coolwarm', 5)
plt.bar(list(ipl['city'].value_counts()[0:5].keys()),list(ipl['city'].valu
e_counts()[0:5]), color = colors(range(5)))
plt.show()
```


100
80
60
40
20
0
Mumbai
Kolkata
Delhi
Bangalore
Hyderabad