



Project Report on **IPL Data Analysis**



Submitted by
Group – 6 Member

Payal Kumari

Ranjit Kumar

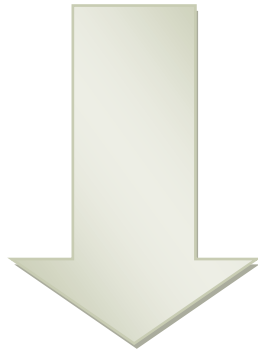
Swarangi Sanjay Raut

Mohammed Mustafa Yusuf Shaikh

Infotact Solution

About Infotact Solution:

Infotact Solution is a leading data-driven company specializing in business intelligence, analytics, and AI-powered insights. We help organizations across industries leverage data to make informed decisions through innovative solutions, including data visualization, predictive modeling, and business optimization.



EDA

Infotact Solution

iplanalysis

February 18, 2025

1 IPL Analysis

```
[3]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
```

```
[2]: pip install scipy
```

Collecting scipyNote: you may need to restart the kernel to use updated packages.

Downloading scipy-1.15.2-cp313-cp313-win_amd64.whl.metadata (60 kB)
Requirement already satisfied: numpy<2.5,>=1.23.5 in
c:\users\windows\appdata\local\programs\python\python313\lib\site-packages (from
scipy) (2.2.0)

Downloading scipy-1.15.2-cp313-cp313-win_amd64.whl (41.0 MB)

```
----- 0.0/41.0 MB ? eta -:-:--
- ----- 1.3/41.0 MB 7.1 MB/s eta 0:00:06
-- ----- 2.6/41.0 MB 6.8 MB/s eta 0:00:06
--- ----- 3.7/41.0 MB 6.1 MB/s eta 0:00:07
----- 5.2/41.0 MB 6.3 MB/s eta 0:00:06
----- 6.8/41.0 MB 6.5 MB/s eta 0:00:06
----- 8.1/41.0 MB 6.5 MB/s eta 0:00:06
----- 9.7/41.0 MB 6.7 MB/s eta 0:00:05
----- 11.5/41.0 MB 6.9 MB/s eta 0:00:05
----- 13.1/41.0 MB 7.0 MB/s eta 0:00:04
----- 14.9/41.0 MB 7.2 MB/s eta 0:00:04
----- 16.8/41.0 MB 7.3 MB/s eta 0:00:04
----- 18.6/41.0 MB 7.4 MB/s eta 0:00:04
----- 20.7/41.0 MB 7.5 MB/s eta 0:00:03
----- 22.5/41.0 MB 7.6 MB/s eta 0:00:03
----- 24.4/41.0 MB 7.7 MB/s eta 0:00:03
----- 26.5/41.0 MB 7.8 MB/s eta 0:00:02
----- 28.6/41.0 MB 7.9 MB/s eta 0:00:02
----- 30.4/41.0 MB 8.0 MB/s eta 0:00:02
----- 32.2/41.0 MB 8.0 MB/s eta 0:00:02
----- 34.6/41.0 MB 8.1 MB/s eta 0:00:01
```

```

----- 36.4/41.0 MB 8.2 MB/s eta 0:00:01
----- 38.3/41.0 MB 8.1 MB/s eta 0:00:01
----- 40.6/41.0 MB 8.2 MB/s eta 0:00:01
----- 40.9/41.0 MB 8.2 MB/s eta 0:00:01
----- 40.9/41.0 MB 8.2 MB/s eta 0:00:01
----- 41.0/41.0 MB 7.5 MB/s eta 0:00:00

```

Installing collected packages: scipy
 Successfully installed scipy-1.15.2

[notice] A new release of pip is available: 24.3.1 -> 25.0.1
 [notice] To update, run: python.exe -m pip install --upgrade pip

1.0.1 Load Dataset

```
[6]: deliveries_path = "E:/Infotact/archive/deliveries.csv"
     matches_path = "E:/Infotact/archive/matches.csv"

     deliveries_df = pd.read_csv(deliveries_path)
     matches_df = pd.read_csv(matches_path)
```

```
[7]: deliveries_df.columns
     deliveries_df.describe()
     deliveries_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 260920 entries, 0 to 260919
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   match_id              260920 non-null  int64
1   inning                260920 non-null  int64
2   batting_team          260920 non-null  object
3   bowling_team          260920 non-null  object
4   over                  260920 non-null  int64
5   ball                  260920 non-null  int64
6   batter                260920 non-null  object
7   bowler                260920 non-null  object
8   non_striker           260920 non-null  object
9   batsman_runs          260920 non-null  int64
10  extra_runs            260920 non-null  int64
11  total_runs            260920 non-null  int64
12  extras_type           14125 non-null   object
13  is_wicket             260920 non-null  int64
14  player_dismissed      12950 non-null   object
15  dismissal_kind        12950 non-null   object
16  fielder               9354 non-null    object
dtypes: int64(8), object(9)
```

memory usage: 33.8+ MB

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

```
[10]: (260920, 17)
```

```
[9]: deliveries_df.head(10)
```

```
[9]:
```

	match_id	inning	batting_team	bowling_team	over	\
0	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	0	
1	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	0	
2	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	0	
3	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	0	
4	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	0	
5	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	0	
6	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	0	
7	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	1	
8	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	1	
9	335982	1	Kolkata Knight Riders	Royal Challengers Bangalore	1	

	ball	batter	bowler	non_striker	batsman_runs	extra_runs	\
0	1	SC Ganguly	P Kumar	BB McCullum	0	1	
1	2	BB McCullum	P Kumar	SC Ganguly	0	0	
2	3	BB McCullum	P Kumar	SC Ganguly	0	1	
3	4	BB McCullum	P Kumar	SC Ganguly	0	0	
4	5	BB McCullum	P Kumar	SC Ganguly	0	0	
5	6	BB McCullum	P Kumar	SC Ganguly	0	0	
6	7	BB McCullum	P Kumar	SC Ganguly	0	1	
7	1	BB McCullum	Z Khan	SC Ganguly	0	0	
8	2	BB McCullum	Z Khan	SC Ganguly	4	0	
9	3	BB McCullum	Z Khan	SC Ganguly	4	0	

	total_runs	extras_type	is_wicket	player_dismissed	dismissal_kind	fielder
0	1	legbyes	0	NaN	NaN	NaN
1	0	NaN	0	NaN	NaN	NaN
2	1	wides	0	NaN	NaN	NaN
3	0	NaN	0	NaN	NaN	NaN
4	0	NaN	0	NaN	NaN	NaN
5	0	NaN	0	NaN	NaN	NaN
6	1	legbyes	0	NaN	NaN	NaN
7	0	NaN	0	NaN	NaN	NaN
8	4	NaN	0	NaN	NaN	NaN
9	4	NaN	0	NaN	NaN	NaN

```
[8]: deliveries_df.tail(5)
```

```
[8]:
```

	match_id	inning	batting_team		bowling_team		over	\
260915	1426312	2	Kolkata Knight Riders	Sunrisers Hyderabad			9	
260916	1426312	2	Kolkata Knight Riders	Sunrisers Hyderabad			9	
260917	1426312	2	Kolkata Knight Riders	Sunrisers Hyderabad			10	
260918	1426312	2	Kolkata Knight Riders	Sunrisers Hyderabad			10	
260919	1426312	2	Kolkata Knight Riders	Sunrisers Hyderabad			10	

	ball	batter	bowler	non_striker	batsman_runs	extra_runs	\
260915	5	SS Iyer	AK Markram	VR Iyer	1	0	
260916	6	VR Iyer	AK Markram	SS Iyer	1	0	
260917	1	VR Iyer	Shahbaz Ahmed	SS Iyer	1	0	
260918	2	SS Iyer	Shahbaz Ahmed	VR Iyer	1	0	
260919	3	VR Iyer	Shahbaz Ahmed	SS Iyer	1	0	

	total_runs	extras_type	is_wicket	player_dismissed	dismissal_kind	\
260915	1	NaN	0	NaN	NaN	
260916	1	NaN	0	NaN	NaN	
260917	1	NaN	0	NaN	NaN	
260918	1	NaN	0	NaN	NaN	
260919	1	NaN	0	NaN	NaN	

	fielder
260915	NaN
260916	NaN
260917	NaN
260918	NaN
260919	NaN

```
[13]: matches_df.columns
matches_df.describe()
matches_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   object
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(1), object(16)
memory usage: 171.2+ KB

```

```
[14]: matches_df.shape
```

```
[14]: (1095, 20)
```

```
[15]: matches_df.head(5)
```

```

[15]:      id  season      city      date match_type player_of_match \
0  335982  2007/08  Bangalore  2008-04-18    League    BB McCullum
1  335983  2007/08  Chandigarh  2008-04-19    League    MEK Hussey
2  335984  2007/08    Delhi  2008-04-19    League    MF Maharoor
3  335985  2007/08    Mumbai  2008-04-20    League    MV Boucher
4  335986  2007/08   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

```
[16]: matches_df.tail(5)
```

```
[16]:
```

	id	season	city	date	match_type	player_of_match	\
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	\
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	toss_winner	\
1090	Punjab Kings	Sunrisers Hyderabad	Punjab Kings	
1091	Sunrisers Hyderabad	Kolkata Knight Riders	Sunrisers Hyderabad	
1092	Royal Challengers Bengaluru	Rajasthan Royals	Rajasthan Royals	
1093	Sunrisers Hyderabad	Rajasthan Royals	Rajasthan Royals	
1094	Sunrisers Hyderabad	Kolkata Knight Riders	Sunrisers Hyderabad	

	toss_decision	winner	result	result_margin	\
1090	bat	Sunrisers Hyderabad	wickets	4.0	
1091	bat	Kolkata Knight Riders	wickets	8.0	
1092	field	Rajasthan Royals	wickets	4.0	
1093	field	Sunrisers Hyderabad	runs	36.0	
1094	bat	Kolkata Knight Riders	wickets	8.0	

	target_runs	target_overs	super_over	method	umpire1	\
1090	215.0	20.0	N	NaN	Nitin Menon	
1091	160.0	20.0	N	NaN	AK Chaudhary	
1092	173.0	20.0	N	NaN	KN Ananthapadmanabhan	
1093	176.0	20.0	N	NaN	Nitin Menon	
1094	114.0	20.0	N	NaN	J Madanagopal	

	umpire2
1090	VK Sharma
1091	R Pandit


```
1092 MV Saidharshan Kumar
1093          VK Sharma
1094          Nitin Menon
```

1.0.2 Missing Values

```
[12]: # Checking for missing values
deliveries_missing = deliveries_df.isnull().sum()
deliveries_missing
```

```
[12]: match_id          0
inning              0
batting_team       0
bowling_team       0
over              0
ball              0
batter            0
bowler            0
non_striker        0
batsman_runs       0
extra_runs         0
total_runs         0
extras_type        246795
is_wicket          0
player_dismissed   247970
dismissal_kind     247970
fielder           251566
dtype: int64
```

```
[17]: matches_missing = matches_df.isnull().sum()
matches_missing
```

```
[17]: id              0
season           0
city            51
date            0
match_type      0
player_of_match  5
venue           0
team1           0
team2           0
toss_winner     0
toss_decision   0
winner          5
result          0
result_margin   19
target_runs     3
```

```

target_overs      3
super_over        0
method            1074
umpire1           0
umpire2           0
dtype: int64

```

```

[18]: # Descriptive statistics
deliveries_desc = deliveries_df.describe()
matches_desc = matches_df.describe()

```

Mean, Median and Mode

Deliveries

```

[19]: # Mean, Median, Mode calculations
def calculate_statistics(df, column):
    return {
        "mean": df[column].mean(),
        "median": df[column].median(),
        "mode": df[column].mode()[0] if not df[column].mode().empty else None
    }

# Key numerical columns for deliveries
deliveries_stats = {col: calculate_statistics(deliveries_df, col) for col in
    ↳ ['batsman_runs', 'total_runs', 'extra_runs']}
deliveries_stats

```

```

[19]: {'batsman_runs': {'mean': np.float64(1.265000766518473),
    'median': np.float64(1.0),
    'mode': np.int64(0)},
    'total_runs': {'mean': np.float64(1.3328069906484745),
    'median': np.float64(1.0),
    'mode': np.int64(1)},
    'extra_runs': {'mean': np.float64(0.06780622413000154),
    'median': np.float64(0.0),
    'mode': np.int64(0)}}

```

Matches

```

[20]: # Mean, Median, Mode calculations
def calculate_statistics(df, column):
    return {
        "mean": df[column].mean(),
        "median": df[column].median(),
        "mode": df[column].mode()[0] if not df[column].mode().empty else None
    }

# Key numerical columns for matches

```

```

matches_stats = {col: calculate_statistics(matches_df, col) for col in
    ↪['result_margin', 'target_runs']}
matches_stats

```

```

[20]: {'result_margin': {'mean': np.float64(17.259293680297397),
    'median': np.float64(8.0),
    'mode': np.float64(6.0)},
    'target_runs': {'mean': np.float64(165.68406593406593),
    'median': np.float64(166.0),
    'mode': np.float64(166.0)}}

```

Visualization

```

[21]: # Data visualization
fig, axes = plt.subplots(2, 2, figsize=(12, 10))

# Distribution of Batsman Runs per Ball
sns.histplot(deliveries_df["batsman_runs"], bins=7, kde=True, ax=axes[0, 0])
axes[0, 0].set_title("Distribution of Batsman Runs per Ball")

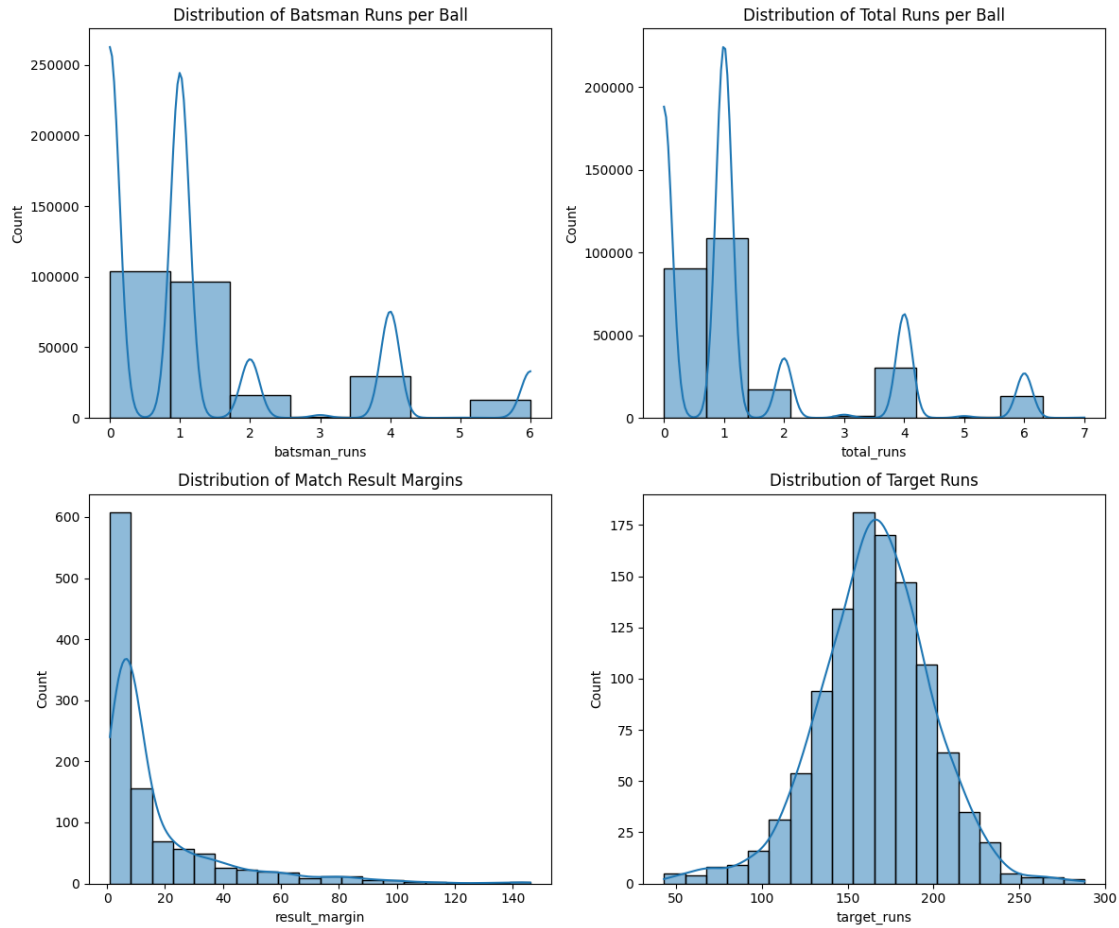
# Distribution of Total Runs per Ball
sns.histplot(deliveries_df["total_runs"], bins=10, kde=True, ax=axes[0, 1])
axes[0, 1].set_title("Distribution of Total Runs per Ball")

# Distribution of Result Margin in Matches
sns.histplot(matches_df["result_margin"].dropna(), bins=20, kde=True,
    ↪ax=axes[1, 0])
axes[1, 0].set_title("Distribution of Match Result Margins")

# Distribution of Target Runs in Matches
sns.histplot(matches_df["target_runs"].dropna(), bins=20, kde=True, ax=axes[1,
    ↪1])
axes[1, 1].set_title("Distribution of Target Runs")

plt.tight_layout()
plt.show()

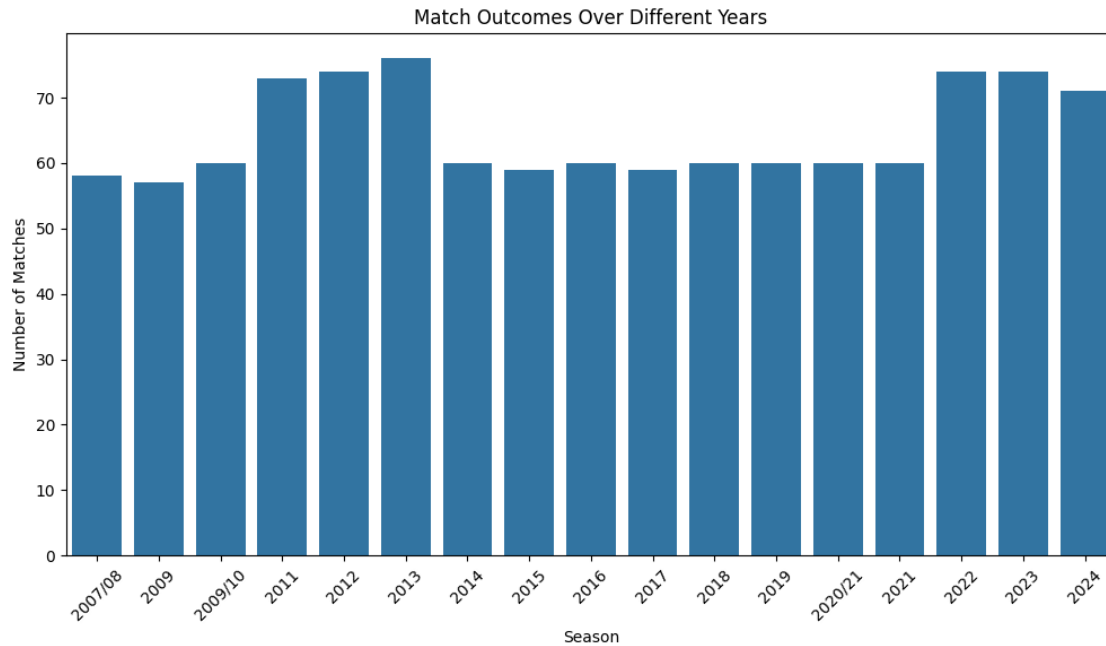
```



1.1 1. Match Outcomes

```
[25]: def match_outcome_analysis(matches):
    plt.figure(figsize=(12, 6))
    sns.countplot(data=matches, x='season', order=sorted(matches['season'].
    ↪unique()))
    plt.title('Match Outcomes Over Different Years')
    plt.xticks(rotation=45)
    plt.xlabel('Season')
    plt.ylabel('Number of Matches')
    plt.show()

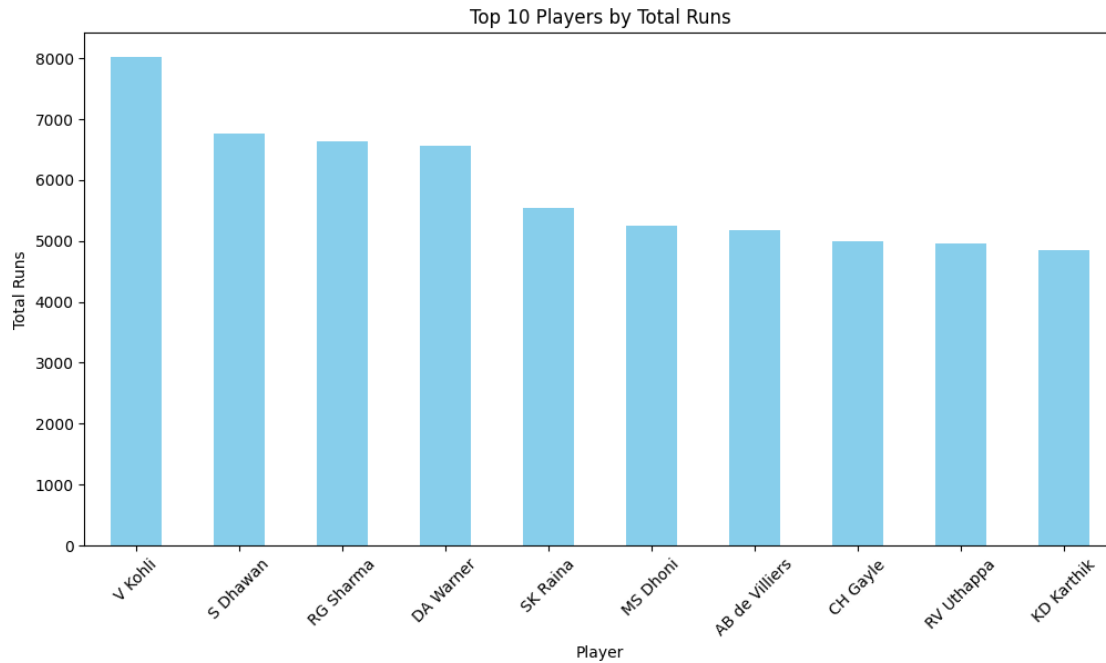
if __name__ == "__main__":
    match_outcome_analysis(matches_df)
```



1.2 2. Player_Performance

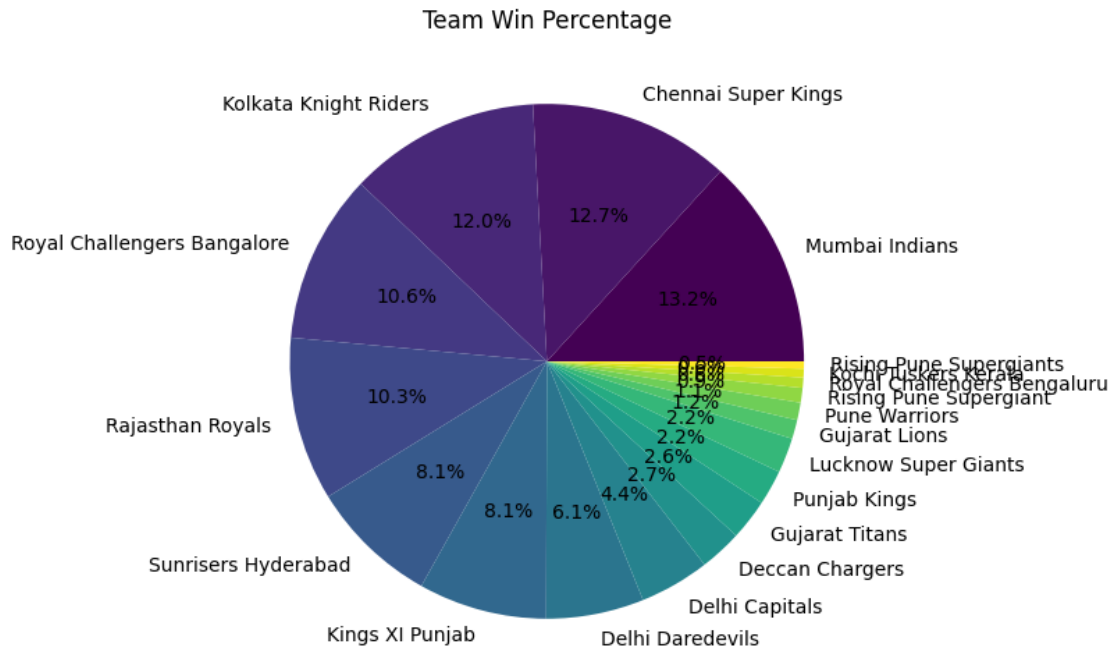
```
[26]: def player_performance(deliveries):
    player_stats = deliveries.groupby('batter')['batsman_runs'].sum().
    ↪sort_values(ascending=False).head(10)
    plt.figure(figsize=(12, 6))
    player_stats.plot(kind='bar', color='skyblue')
    plt.title('Top 10 Players by Total Runs')
    plt.xlabel('Player')
    plt.ylabel('Total Runs')
    plt.xticks(rotation=45)
    plt.show()

if __name__ == "__main__":
    player_performance(deliveries_df)
```



1.3 3. Team_Comparison

```
[27]: def team_comparison(matches):  
    team_wins = matches['winner'].value_counts()  
    plt.figure(figsize=(12, 6))  
    team_wins.plot(kind='pie', autopct='%1.1f%%', colormap='viridis')  
    plt.title('Team Win Percentage')  
    plt.ylabel('')  
    plt.show()  
  
if __name__ == "__main__":  
    team_comparison(matches_df)
```



1.4 4. Venue_performance

```
[28]: def venue_performance(matches):
plt.figure(figsize=(12, 6))
venue_wins = matches['venue'].value_counts().head(10)
sns.barplot(x=venue_wins.values, y=venue_wins.index, palette='coolwarm')
plt.title('Top 10 Venues by Matches Hosted')
plt.xlabel('Number of Matches')
plt.ylabel('Venue')
plt.show()

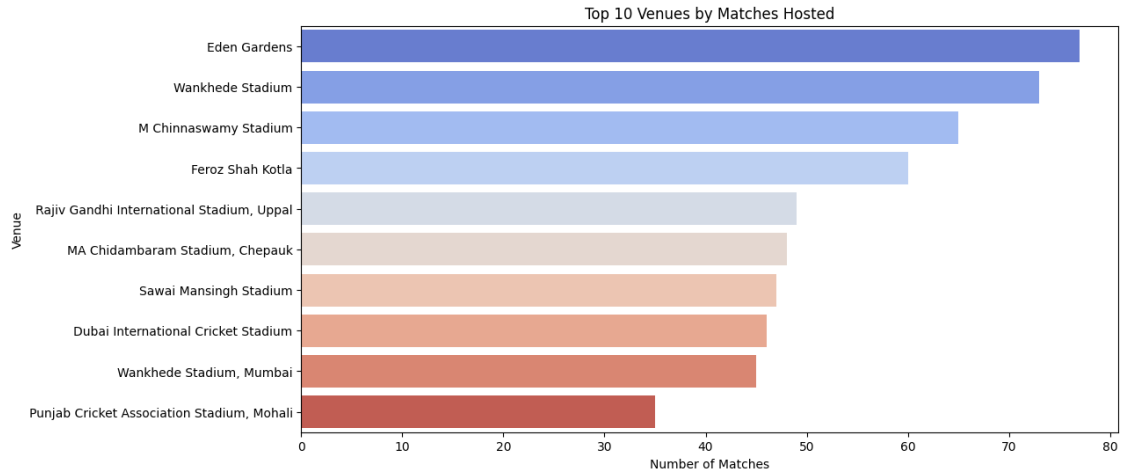
if __name__ == "__main__":
venue_performance(matches_df)
```

C:\Users\Windows\AppData\Local\Temp\ipykernel_11952\4238959835.py:4:

FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

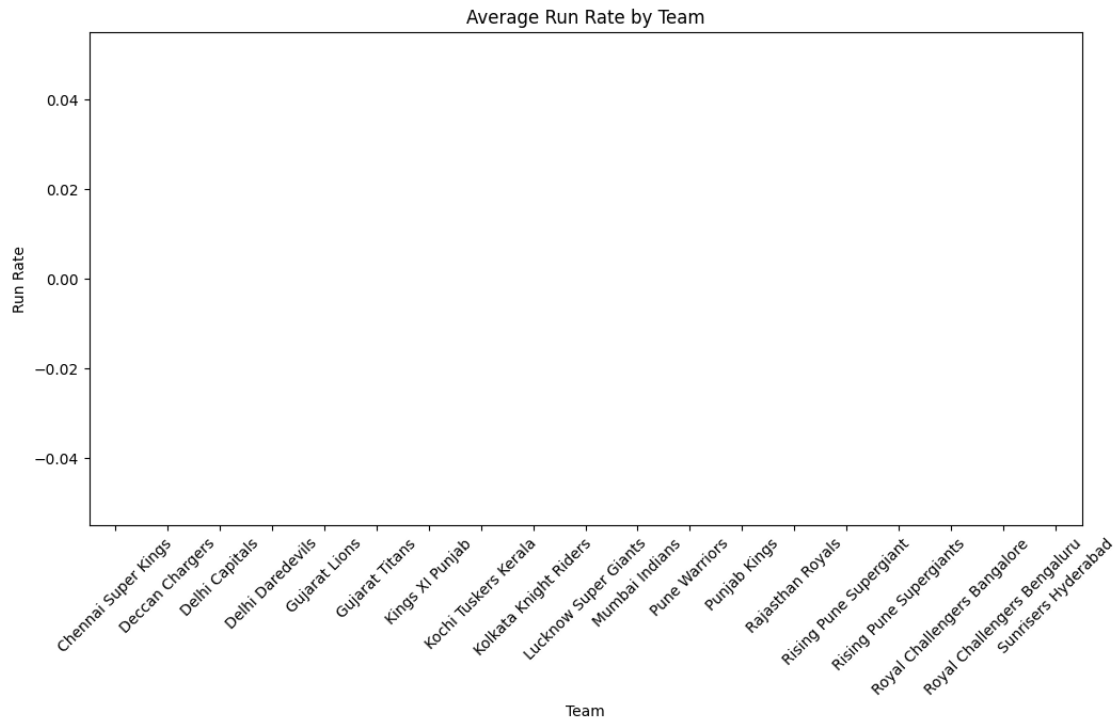
```
sns.barplot(x=venue_wins.values, y=venue_wins.index, palette='coolwarm')
```



1.5 5. Run_Rate_Analysis

```
[29]: def run_rate_analysis(deliveries):
    deliveries['run_rate'] = deliveries['total_runs'] / deliveries['over']
    avg_run_rates = deliveries.groupby('batting_team')['run_rate'].mean().
    ↪sort_values(ascending=False)
    plt.figure(figsize=(12, 6))
    avg_run_rates.plot(kind='bar', color='green')
    plt.title('Average Run Rate by Team')
    plt.xlabel('Team')
    plt.ylabel('Run Rate')
    plt.xticks(rotation=45)
    plt.show()

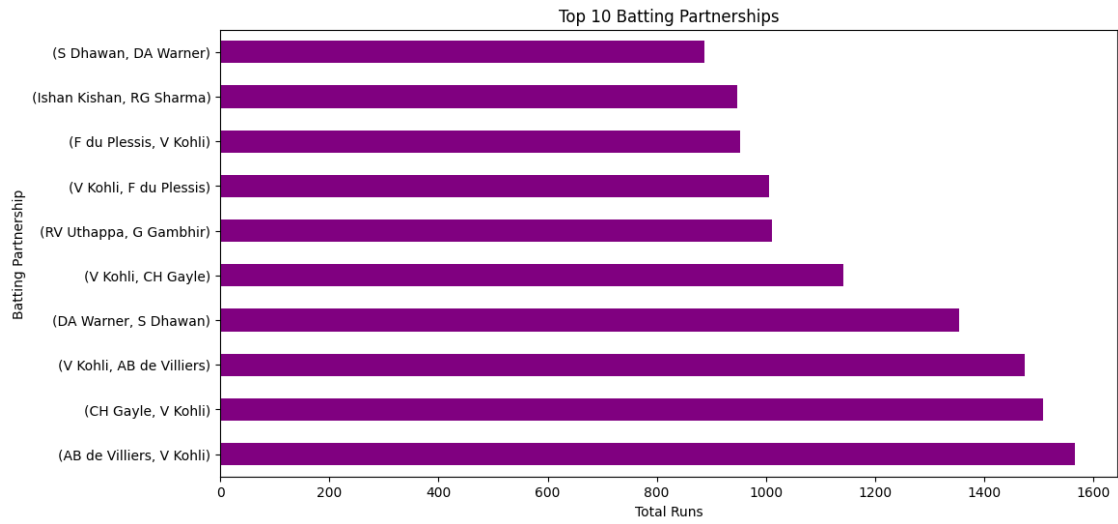
if __name__ == "__main__":
    run_rate_analysis(deliveries_df)
```

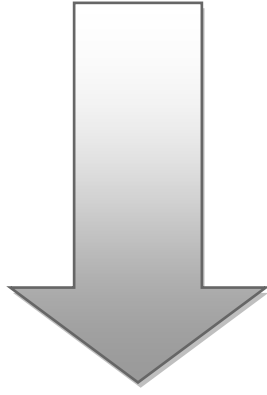



1.6 6. Best_Batting_Partnerships

```
[30]: def best_batting_partnerships(deliveries):
    partnerships = deliveries.groupby(['batter', '
↪ 'non_striker'])['batsman_runs'].sum().sort_values(ascending=False).head(10)
    plt.figure(figsize=(12, 6))
    partnerships.plot(kind='barh', color='purple')
    plt.title('Top 10 Batting Partnerships')
    plt.xlabel('Total Runs')
    plt.ylabel('Batting Partnership')
    plt.show()

if __name__ == "__main__":
    best_batting_partnerships(deliveries_df)
```





Dashboard

Infotact Solution



IPL Data Analysis

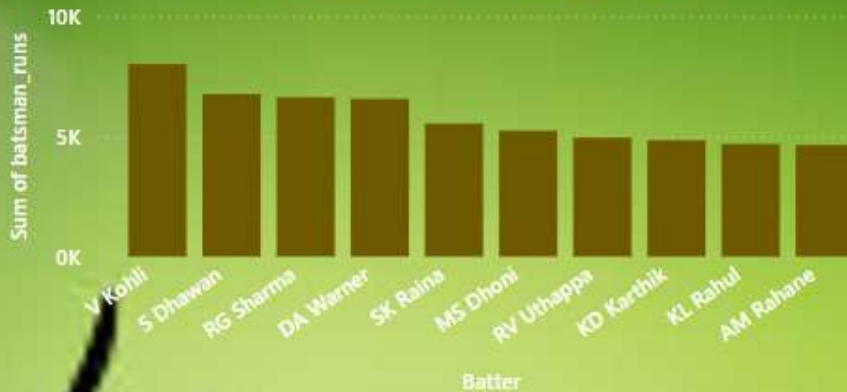




IPL Cricket Deliveries Data



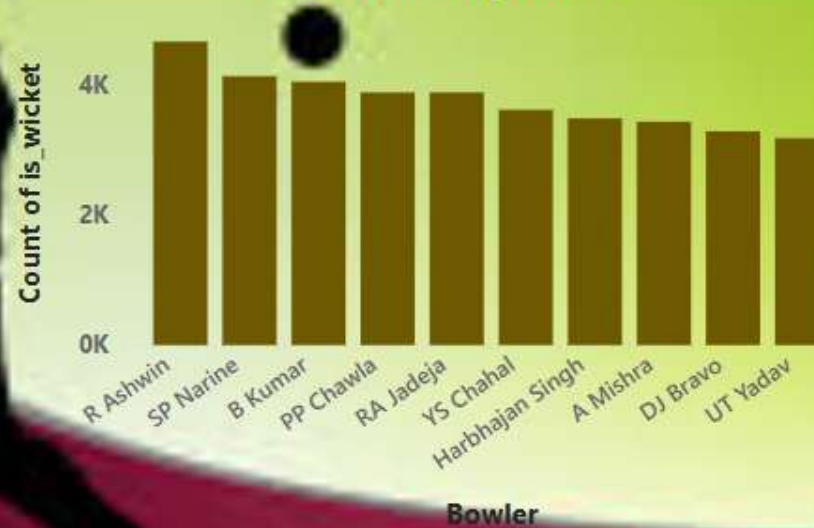
Top batsmen by Runs Scored



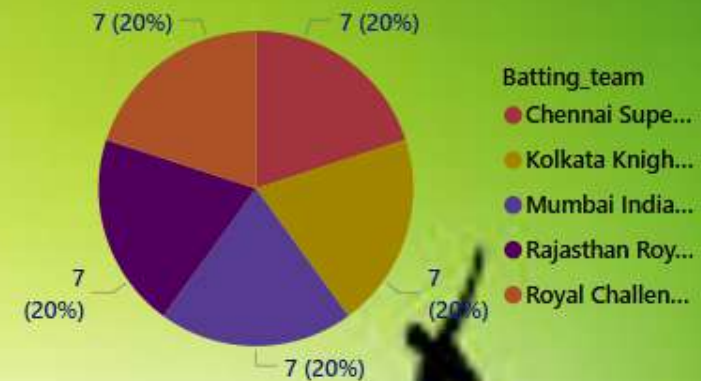
Runs Scored per over



Wicket taken by bowler



Count of batsman_runs by batting_team



330K

Sum of batsman_r...

348K

Sum of total_runs

13K

Sum of is_wicket

387K

Sum of inning

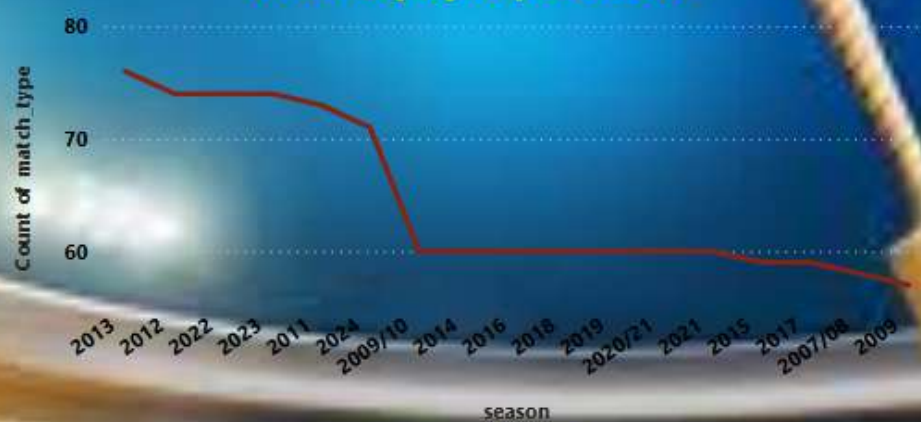
2M

Sum of over

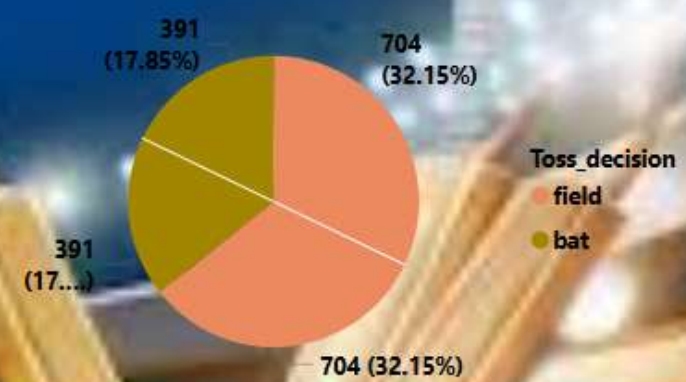


IPL Matches

Matches played per Season



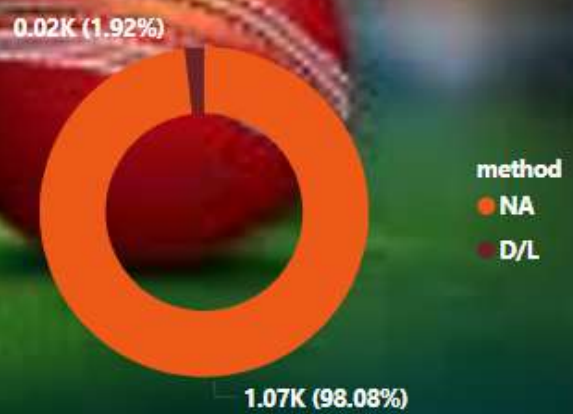
Toss Decision Analysis



Top Teams by Wins



Count of match_type by method



20

Count of winner

8

Count of match_type

21.58K

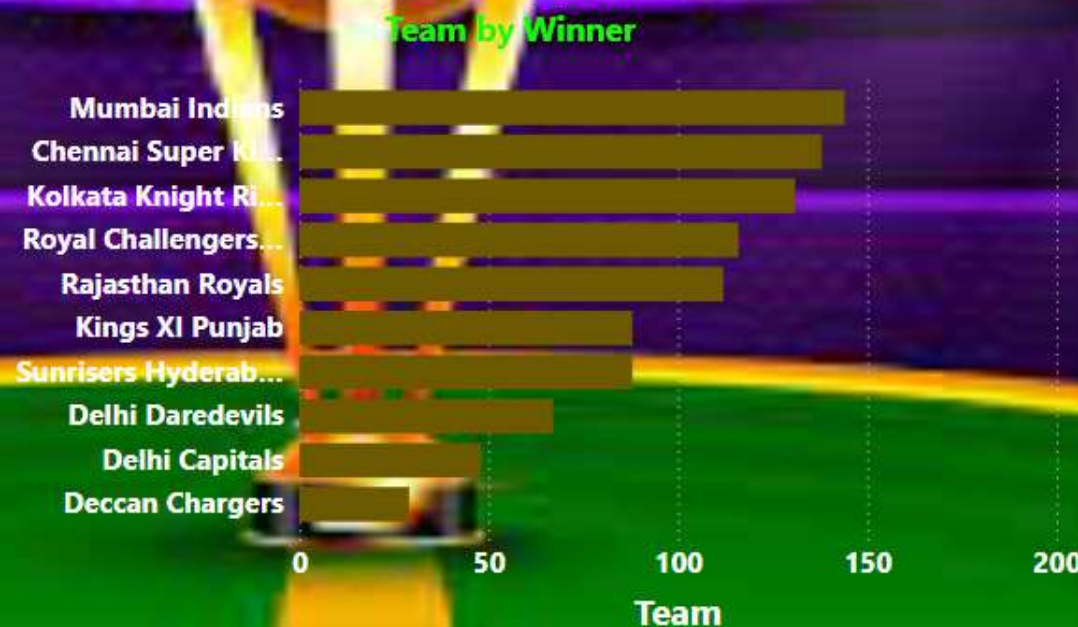
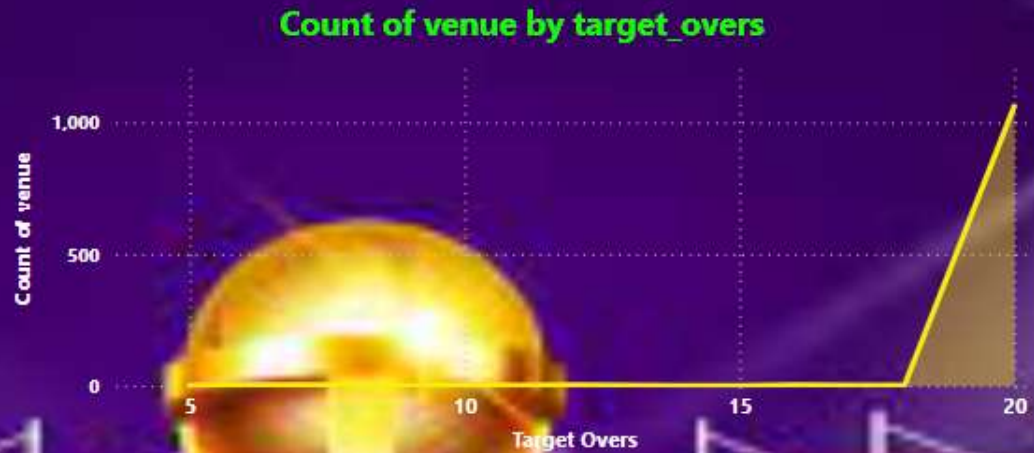
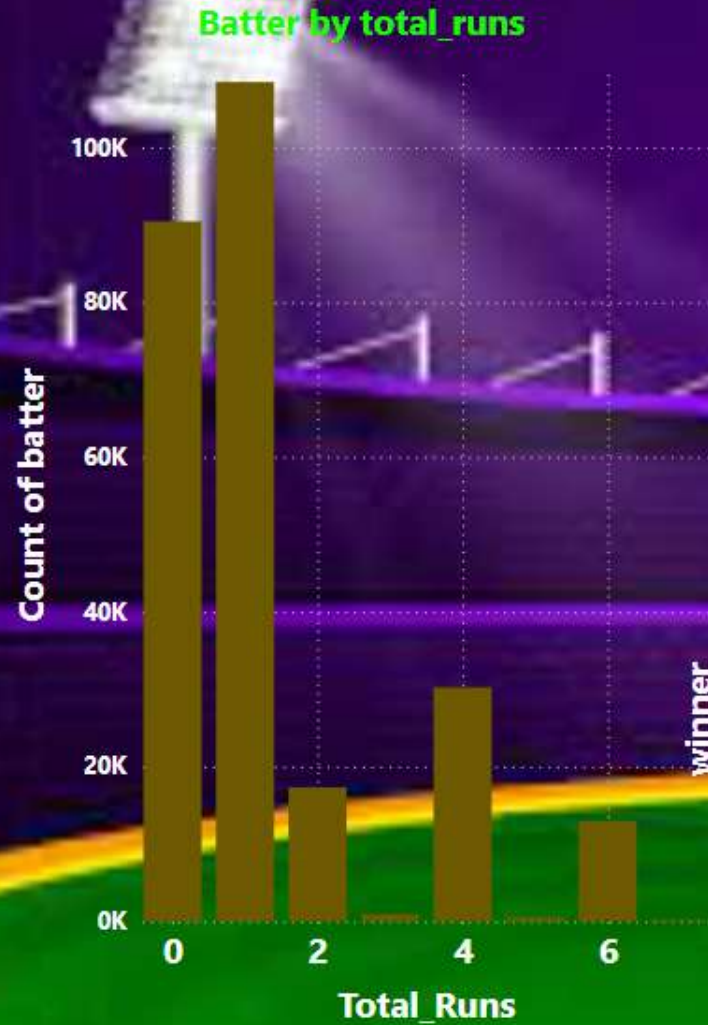
Sum of target_overs

181K

Sum of target_runs



IPL Insights



20
Count of winner

2M
Sum of over

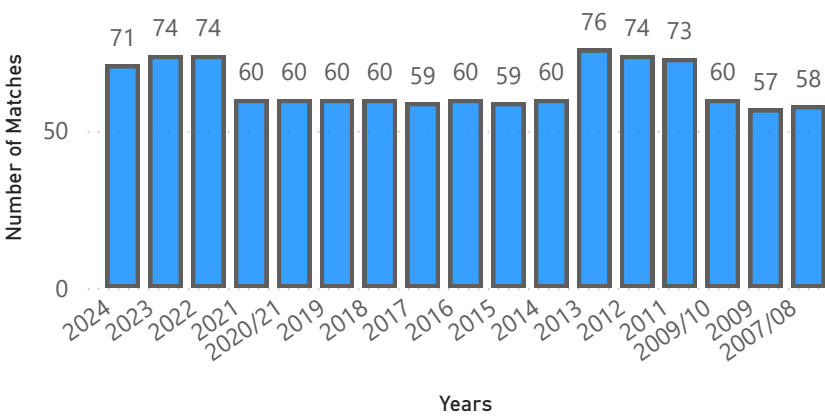
530
Count of bowler

18K
Sum of extra_runs

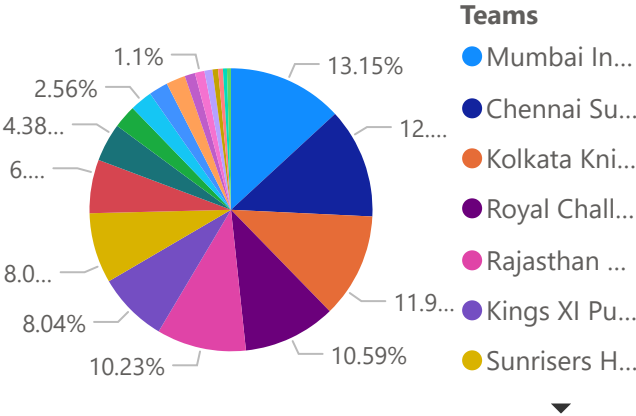
1092
Count of target_overs

IPL ANALYSIS

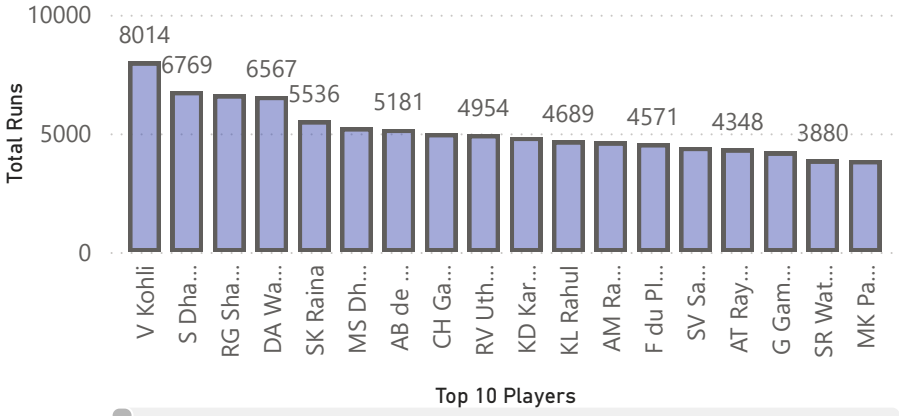
Matches Outcome Over Different Years



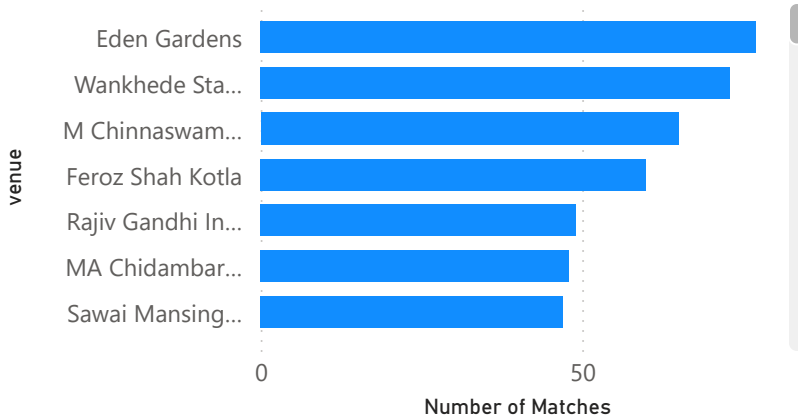
Team Win Percentage



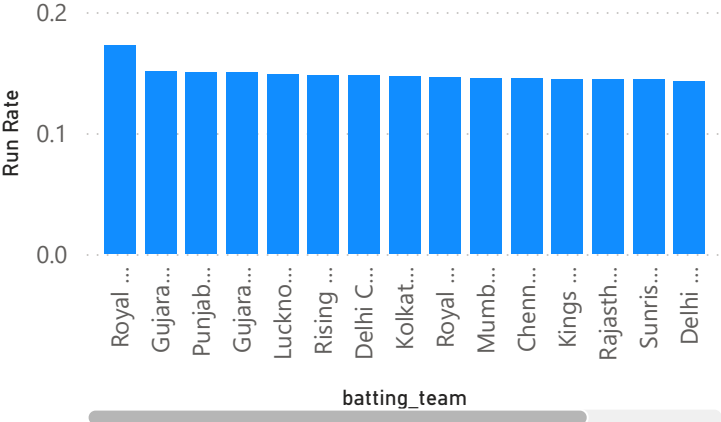
Top 10 Players



Top 10 Venues



Top 10 Batting_teams



Top 10 Batting Partnerships

