

In [194... `import pandas as pd`

Data Exploration

Datasets

we have 6 csv file:

- dim_match_summary
- dim_player_image
- dim_player
- dim_teams
- fact_batting_summary
- fact_bowling_summary

Read match summary

In [195... `df_match_summary = pd.read_csv("D:\PowerBI Works\IPL-Analysis\C10_Input_Files\datas`

Explore match data

In [196... `df_match_summary.head()`

Out[196]:

	team1	team2	winner	margin	matchDate	match_id
0	Super Kings	KKR	Super Kings	27 runs	Oct 15, 2021	T203817
1	Capitals	KKR	KKR	3 wickets	Oct 13, 2021	T206442
2	RCB	KKR	KKR	4 wickets	Oct 11, 2021	T208597
3	Capitals	Super Kings	Super Kings	4 wickets	Oct 10, 2021	T201819
4	Capitals	RCB	RCB	7 wickets	Oct 8, 2021	T205942

In [197... `# Rows and Columns`

`df_match_summary.shape`

Out[197]: (206, 6)

In [198... `# Matches Played`

`df_match_summary.match_id.count()`

Out[198]: 206

Read the rest of the files

In [199... `df_teams = pd.read_csv("D:\PowerBI Works\IPL-Analysis\C10_Input_Files\data`
`df_bowling_summary= pd.read_csv(r"D:\PowerBI Works\IPL-Analysis\C10_Input_Files\dat`
`df_batting_summary= pd.read_csv(r"D:\PowerBI Works\IPL-Analysis\C10_Input_Files\dat`
`df_players = pd.read_csv("D:\PowerBI Works\IPL-Analysis\C10_Input_Files\data`
`df_players_profile= pd.read_csv("D:\PowerBI Works\IPL-Analysis\C10_Input_Files\data`

Explore Teams data

In [200... `df_teams.shape`

Out[200]: (10, 2)

In [201... `df_teams.team.unique()`

Out[201]: array(['Super Kings', 'KKR', 'Capitals', 'RCB', 'Mumbai', 'Royals',
 'Punjab Kings', 'Titans', 'Super Giants', 'Sunrisers'],
 dtype=object)

In [202... `df_teams.team.value_counts()`

Out[202]: Super Kings 1
KKR 1
Capitals 1
RCB 1
Mumbai 1
Royals 1
Punjab Kings 1
Titans 1
Super Giants 1
Sunrisers 1
Name: team, dtype: int64

Explore Batting Data

In [203... `df_batting_summary`

Out[203]:

	match_id	match	teamInnings	battingPos	batsmanName	out/not_out	runs	balls	4s	6s
0	T203817	Super Kings Vs KKR	Super Kings	1	RuturajGaikwad	out	32	27	3	1
1	T203817	Super Kings Vs KKR	Super Kings	2	FafduPlessis	out	86	59	7	3
2	T203817	Super Kings Vs KKR	Super Kings	3	RobinUthappa	out	31	15	0	3
3	T203817	Super Kings Vs KKR	Super Kings	4	MoeenAli	not_out	37	20	2	3
4	T203817	Super Kings Vs KKR	KKR	1	ShubmanGill	out	51	43	6	0
...
3263	T201566	Super Kings Vs Titans	Titans	3	SaiSudharsan	out	22	17	3	0
3264	T201566	Super Kings Vs Titans	Titans	4	HardikPandya	out	8	11	0	0
3265	T201566	Super Kings Vs Titans	Titans	5	VijayShankar	out	27	21	2	1
3266	T201566	Super Kings Vs Titans	Titans	6	RahulTewatia	not_out	15	14	1	1
3267	T201566	Super Kings Vs Titans	Titans	7	RashidKhan	not_out	10	3	1	1

3268 rows × 11 columns

◀

▶

In [204...

Rows and Columns

df_batting_summary.shape

Out[204]:

(3268, 11)

In [205...

Total number of batsman

```
len(df_batting_summary.drop_duplicates(subset=['batsmanName']))
```

Out[205]: 264

In [206... *# Batsman performance*

```
df_batting_summary.batsmanName.value_counts()
```

Out[206]:

ShubmanGill	50
FafduPlessis	46
RuturajGaikwad	45
ViratKohli	45
SanjuSamson	45
..	..
JoeRoot	1
KMArif	1
ChrisWoakes	1
PraveenDubey	1
ShahbazNadeem	1

Name: batsmanName, Length: 264, dtype: int64

Explore Bowling Data

In [207... df_bowling_summary

Out[207]:

	match_id	match	bowlingTeam	bowlerName	overs	maiden	runs	wickets	economy
0	T203817	Super Kings Vs KKR	KKR	ShakibAlHasan	3.0	0	33	0	11.0
1	T203817	Super Kings Vs KKR	KKR	ShivamMavi	4.0	0	32	1	8.0
2	T203817	Super Kings Vs KKR	KKR	LockieFerguson	4.0	0	56	0	14.0
3	T203817	Super Kings Vs KKR	KKR	VarunChakravarthy	4.0	0	38	0	9.5
4	T203817	Super Kings Vs KKR	KKR	SunilNarine	4.0	0	26	2	6.5
...
2431	T201566	Super Kings Vs Titans	Super Kings	DeepakChahar	4.0	0	29	0	7.25
2432	T201566	Super Kings Vs Titans	Super Kings	TusharDeshpande	3.2	0	51	1	15.94
2433	T201566	Super Kings Vs Titans	Super Kings	RajvardhanHangargekar	4.0	0	36	3	9.0
2434	T201566	Super Kings Vs Titans	Super Kings	MitchellSantner	4.0	0	32	0	8.0
2435	T201566	Super Kings Vs Titans	Super Kings	RavindraJadeja	4.0	0	28	1	7.0

2436 rows × 14 columns

◀

▶

In [208...

Total rows and columns

df_bowling_summary.shape

Out[208]:

(2436, 14)

In [209...

Total Bowling Teams

```
df_bowling_summary.bowlingTeam.unique()
```

Out[209]: array(['KKR', 'Super Kings', 'Capitals', 'RCB', 'Sunrisers', 'Mumbai', 'Royals', 'Punjab Kings', 'Titans', 'Super Giants'], dtype=object)

```
In [210... #Bowlers count  
  
len(df_bowling_summary.drop_duplicates(subset=['bowlerName']))
```

Out[210]: 202

Explore Players Data

```
In [211... df_players
```

Out[211]:

	name	team	battingStyle	bowlingStyle	playingRole
0	RuturajGaikwad	Super Kings	Right hand Bat	Right arm Offbreak	Batter
1	FafduPlessis	Super Kings	Right hand Bat	Legbreak	Middle order Batter
2	RobinUthappa	Super Kings	Right hand Bat	Right arm Medium	Batter
3	MoeenAli	Super Kings	Left hand Bat	Right arm Offbreak	Batting Allrounder
4	ShubmanGill	KKR	Right hand Bat	Right arm Offbreak	Opening Batter
...
287	AbishekPorel	Capitals	Left hand Bat	Not Available	Wicketkeeper Batter
288	SisandaMagala	Super Kings	Right hand Bat	Right arm Fast medium	Bowler
289	MohitRathee	Punjab Kings	Right hand Bat	Legbreak	Allrounder
290	RajvardhanHangargekar	Super Kings	Right hand Bat	Right arm Fast medium	Bowling Allrounder
291	ReeceTopley	RCB	Right hand Bat	Left arm Fast medium	Bowler

292 rows x 5 columns

```
In [212... df_players.shape
```

Out[212]: (292, 5)

```
In [213... # Totla Players  
len(df_players.drop_duplicates(['name']))
```

Out[213]: 292

Explore Players Profile Data

```
In [214... df_players_profile
```

Out[214]:

	name	team	image
0	RuturajGaikwad	Super Kings	https://bcciplayerimages.s3.ap-south-1.amazona...
1	FafduPlessis	Super Kings	https://parimatchnews.com/wp-content/uploads/2...
2	RobinUthappa	Super Kings	https://bcciplayerimages.s3.ap-south-1.amazona...
3	MoeenAli	Super Kings	https://bcciplayerimages.s3.ap-south-1.amazona...
4	ShubmanGill	KKR	https://parimatchnews.com/wp-content/uploads/2...
...
287	AbishekPorel	Capitals	https://bcciplayerimages.s3.ap-south-1.amazona...
288	SisandaMagala	Super Kings	https://i0.wp.com/crictoday.com/wp-content/upl...
289	MohitRathee	Punjab Kings	https://www.punjabkingsipl.in/static-assets/im...
290	RajvardhanHangargekar	Super Kings	https://iplbetonline.b-cdn.net/wp-content/uplo...
291	ReeceTopley	RCB	https://indianden.in/wp-content/uploads/2023/0...

292 rows × 3 columns

In [215...

```

# Show Players with image

from IPython.display import display, Image, HTML

def display_images_with_names(df_players_profile):
    # Start HTML table
    html_content = "<table>"
    html_content += "<tr><th>Image</th><th>Name</th></tr>"

    for index, row in df_players_profile.head().iterrows():
        name = row['name']
        image_url = row['image']
        # Check if the image_url is not NaN (missing) and is a valid string
        if isinstance(image_url, str):
            try:
                # Add row to table with image and name
                html_content += f"<tr><td><img src='{image_url}' width='150' height="
            except Exception as e:
                print(f"Error displaying image for {name}: {e}")
                continue
        else:
            print(f"Invalid image URL for {name}, skipping.")
            continue

    # End HTML table
    html_content += "</table>"

    # Display HTML content
    display(HTML(html_content))

# Display images with names in separate columns
display_images_with_names(df_players_profile)

```

Image

Name



RuturajGaikwad



FafduPlessis



RobinUthappa



MoeenAli



ShubmanGill

Data Cleaning and Transformation

In [216...

```
# Handling missing values  
df_match_summary.isnull().head(5)  
  
# There is no duplicate, null and NA values
```


Out[216]:

	team1	team2	winner	margin	matchDate	match_id
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False

In [217... *# Before change dtype*

```
df_match_summary.dtypes
```

Out[217]:

```
team1      object
team2      object
winner     object
margin     object
matchDate  object
match_id   object
dtype: object
```

In [218... `df_match_summary['matchDate']=pd.to_datetime(df_match_summary['matchDate'])`

In [219... *# transform date '2029-05-28' into 2023-05-28*

```
import pandas as pd

# Set display options to show all rows
pd.set_option('display.max_rows', None)

# Display the entire DataFrame
```

In [220... `df_match_summary.loc[133,'matchDate']=pd.to_datetime('2023-05-28')`

```
pd.set_option('display.max_rows', None)

# Display the entire DataFrame

# Reset display options to default values
pd.reset_option('display.max_rows')
```

In [221... *# Create 'Year' column from matchDate*

```
df_match_summary['Year']=df_match_summary['matchDate'].dt.year
df_match_summary
```

Out[221]:

	team1	team2	winner	margin	matchDate	match_id	Year
0	Super Kings	KKR	Super Kings	27 runs	2021-10-15	T203817	2021
1	Capitals	KKR	KKR	3 wickets	2021-10-13	T206442	2021
2	RCB	KKR	KKR	4 wickets	2021-10-11	T208597	2021
3	Capitals	Super Kings	Super Kings	4 wickets	2021-10-10	T201819	2021
4	Capitals	RCB	RCB	7 wickets	2021-10-08	T205942	2021
...
201	Mumbai	RCB	RCB	8 wickets	2023-04-02	T202485	2023
202	Royals	Sunrisers	Royals	72 runs	2023-04-02	T209117	2023
203	Super Giants	Capitals	Super Giants	50 runs	2023-04-01	T208512	2023
204	Punjab Kings	KKR	Punjab Kings	7 runs	2023-04-01	T202853	2023
205	Super Kings	Titans	Titans	5 wickets	2023-03-31	T201566	2023

206 rows × 7 columns

In [222...

Before change dtype

df_match_summary.dtypes

Out[222]:

team1object
team2object
winnerobject
marginobject
matchDatedatetime64[ns]
match_idobject
Yearint64
dtype: object

Check rest of file data types

In [223...

df_batting_summary.dtypes

Out[223]:

match_idobject
matchobject
teamInningsobject
battingPosint64
batsmanNameobject
out/not_outobject
runsint64
ballsint64
4sint64
6sint64
SRobject
dtype: object

In [224...

df_bowling_summary.dtypes

```
Out[224]: match_id      object
          match        object
          bowlingTeam  object
          bowlerName   object
          overs        float64
          maiden       int64
          runs         int64
          wickets      int64
          economy      float64
          0s           int64
          4s           int64
          6s           int64
          wides        int64
          noBalls      int64
          dtype: object
```

```
In [225... df_players.dtypes
```

```
Out[225]: name      object
          team      object
          battingStyle object
          bowlingStyle object
          playingRole object
          dtype: object
```

```
In [226... df_players_profile.dtypes
```

```
Out[226]: name      object
          team      object
          image      object
          dtype: object
```

Insights Generation

1. How many matches played by each team?

```
In [227... import matplotlib.pyplot as plt
```

```
In [228... merged_team1 = pd.merge(df_teams, df_match_summary, left_on='team', right_on='team1')
          merged_team2 = pd.merge(df_teams, df_match_summary, left_on='team', right_on='team2')
```

```
In [229... merged_team1.drop_duplicates(subset=['match_id'], inplace=True)
          merged_team1
```

Out[229]:

	team	logo	team1	team2	winner	margin	matchDate	match_id
0	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	KKR	Super Kings	27 runs	2021-10-15	T203817
1	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	Punjab Kings	Punjab Kings	6 wickets	2021-10-07	T201866
2	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	Capitals	Capitals	3 wickets	2021-10-04	T207682
3	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	Royals	Royals	7 wickets	2021-10-02	T205672
4	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	Mumbai	Super Kings	20 runs	2021-09-19	T201256
...
201	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Sunrisers	Super Giants	Super Giants	7 wickets	2023-05-13	T204737
202	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Sunrisers	Capitals	Sunrisers	9 runs	2023-04-29	T207344
203	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Sunrisers	Super Kings	Super Kings	7 wickets	2023-04-21	T208568
204	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Sunrisers	KKR	Sunrisers	23 runs	2023-04-14	T209839
205	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Sunrisers	Super Giants	Super Giants	5 wickets	2023-04-07	T209590

206 rows × 9 columns

◀

▶

In [230...merged_team2.drop_duplicates(subset=['match_id'], inplace=True)
merged_team2

Out[230]:

	team	logo	team1	team2	winner	margin	matchDate	match_id
0	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Capitals	Super Kings	Super Kings	4 wickets	2021-10-10	T201819
1	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Sunrisers	Super Kings	Super Kings	6 wickets	2021-09-30	T20306
2	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	KKR	Super Kings	Super Kings	2 wickets	2021-09-26	T202406
3	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	RCB	Super Kings	Super Kings	6 wickets	2021-09-24	T207776
4	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Sunrisers	Super Kings	Super Kings	7 wickets	2021-04-28	T205094
...
201	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	KKR	Sunrisers	KKR	5 runs	2023-05-04	T20861
202	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Capitals	Sunrisers	Capitals	7 runs	2023-04-24	T206316
203	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Mumbai	Sunrisers	Mumbai	14 runs	2023-04-18	T20383
204	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Punjab Kings	Sunrisers	Sunrisers	8 wickets	2023-04-09	T206444
205	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Royals	Sunrisers	Royals	72 runs	2023-04-02	T20911

206 rows × 9 columns

In [231]:

```
matches_team1_count = merged_team1['team'].value_counts()
matches_team2_count = merged_team2['team'].value_counts()
all_matches = pd.concat([merged_team1, merged_team2])
all_matches
```

Out[231]:

	team	logo	team1	team2	winner	margin	matchDate	match_id
0	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	KKR	Super Kings	27 runs	2021-10-15	T203817
1	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	Punjab Kings	Punjab Kings	6 wickets	2021-10-07	T201866
2	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	Capitals	Capitals	3 wickets	2021-10-04	T207682
3	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	Royals	Royals	7 wickets	2021-10-02	T205672
4	Super Kings	https://encrypted-tbn0.gstatic.com/images?q=tb...	Super Kings	Mumbai	Super Kings	20 runs	2021-09-19	T201256
...
201	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	KKR	Sunrisers	KKR	5 runs	2023-05-04	T208611
202	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Capitals	Sunrisers	Capitals	7 runs	2023-04-24	T206316
203	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Mumbai	Sunrisers	Mumbai	14 runs	2023-04-18	T203831
204	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Punjab Kings	Sunrisers	Sunrisers	8 wickets	2023-04-09	T206444
205	Sunrisers	https://encrypted-tbn0.gstatic.com/images?q=tb...	Royals	Sunrisers	Royals	72 runs	2023-04-02	T209117

412 rows × 9 columns

◀

▶

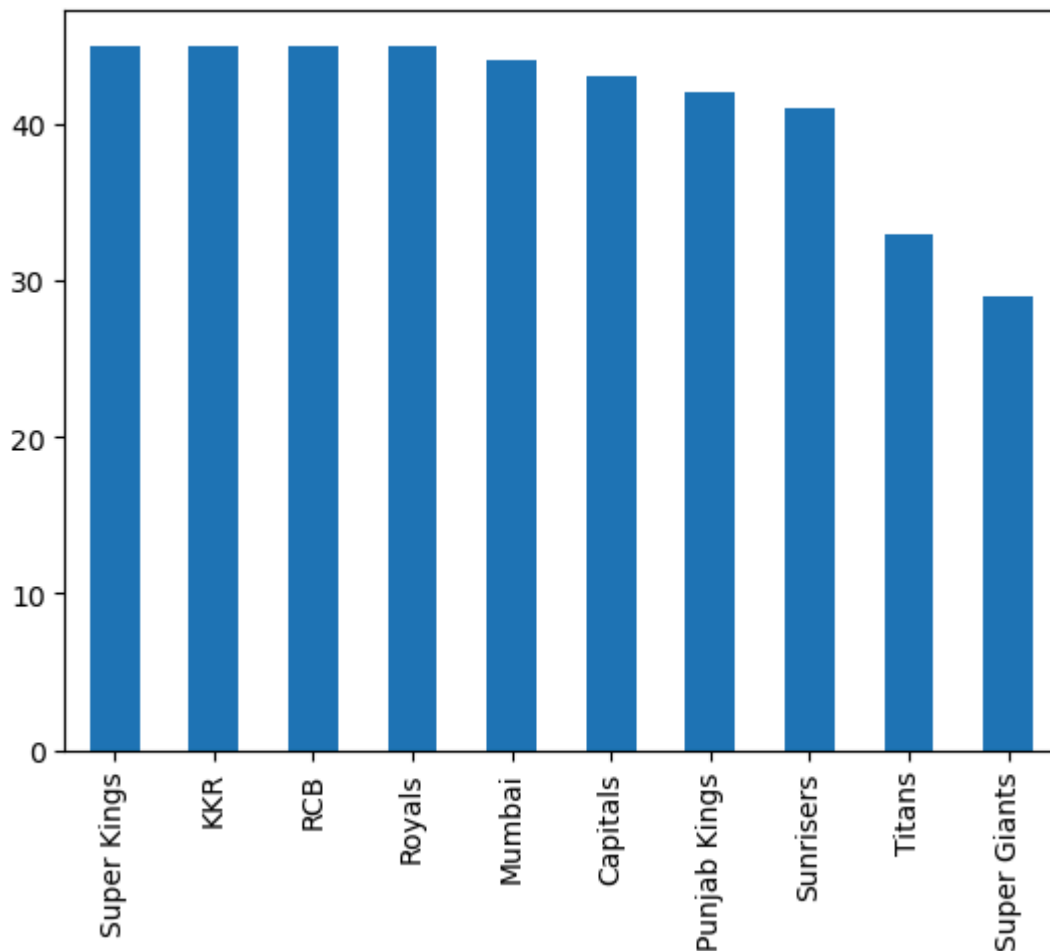
In [232...

```
# Total matches each team played overall

total_matches_count = all_matches['team'].value_counts()
total_matches_count.columns = ['team', 'total_matches']

total_matches_count.plot(kind="bar")
```

Out[232]: <Axes: >



In [233...] total_matches_count

Out[233]:

Super Kings	45
KKR	45
RCB	45
Royals	45
Mumbai	44
Capitals	43
Punjab Kings	42
Sunrisers	41
Titans	33
Super Giants	29

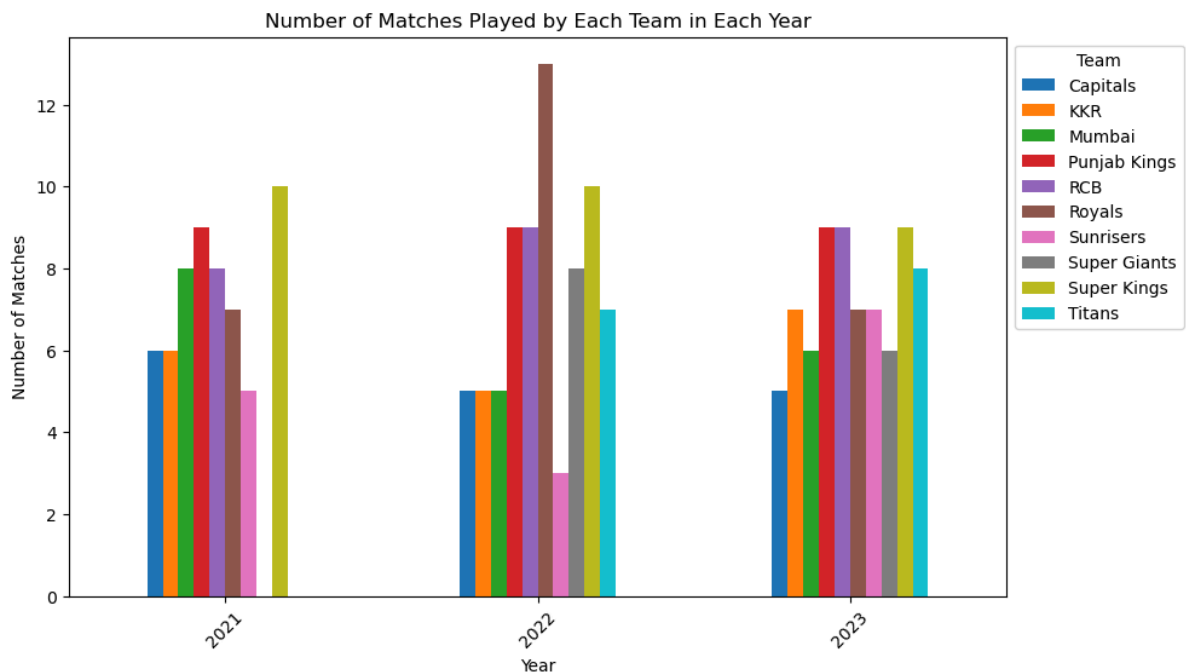
Name: team, dtype: int64

```
In [234...] import matplotlib.pyplot as plt

# Group the DataFrame by 'Year' and 'team' columns, then count the occurrences of each team
team_matches_by_year = merged_team1.groupby(['Year', 'team']).size().unstack()

# Plotting
team_matches_by_year.plot(kind='bar', figsize=(10, 6))
plt.title('Number of Matches Played by Each Team in Each Year')
plt.xlabel('Year')
plt.ylabel('Number of Matches')
plt.xticks(rotation=45)
plt.legend(title='Team', bbox_to_anchor=(1, 1), loc='upper left')

# Show the plot
plt.show()
```



2. How many Matches won?

```
In [235... # Total matches won

merged_winner_team=pd.merge(df_teams,df_match_summary,left_on="team",right_on="winr
total_matches_won=merged_winner_team.team.value_counts()
total_matches_won
```

```
Out[235]: Super Kings      25
RCB          25
Titans       23
Royals       22
KKR          21
Capitals     21
Mumbai       20
Punjab Kings 19
Super Giants 17
Sunrisers    13
Name: team, dtype: int64
```

```
In [236... # Total matches Loss

total_matches_loss=total_matches_count-total_matches_won
total_matches_loss
```

```
Out[236]: Capitals      22
KKR          24
Mumbai       24
Punjab Kings 23
RCB          20
Royals       23
Sunrisers    28
Super Giants 12
Super Kings  20
Titans       10
Name: team, dtype: int64
```

```
In [237... # win %

win_pct =round((total_matches_won/total_matches_count)*100,2)

# concatenating all three match stats
```



```
result = pd.concat([total_matches_count,total_matches_won,total_matches_loss,win_per]
result.columns=['total_matches','total_matches_won','total_matches_loss','win%']
team_stats=result.sort_values('win%',ascending=False)
team_stats
```

Out[237]:

	total_matches	total_matches_won	total_matches_loss	win%
Titans	33	23	10	69.70
Super Giants	29	17	12	58.62
Super Kings	45	25	20	55.56
RCB	45	25	20	55.56
Royals	45	22	23	48.89
Capitals	43	21	22	48.84
KKR	45	21	24	46.67
Mumbai	44	20	24	45.45
Punjab Kings	42	19	23	45.24
Sunrisers	41	13	28	31.71

3. Top 10 batsmen based on past 3 years total runs scored.

In [238...]

```
# Total runs scored

batsman_runs_scored=df_batting_summary.groupby('batsmanName')['runs'].sum()
batsman_runs_scored_sorted=batsman_runs_scored.sort_values(ascending=False)
batsman_runs_scored_sorted.head(10)
```

Out[238]:

```
batsmanName
ShubmanGill      1851
FafduPlessis     1831
RuturajGaikwad   1593
KL Rahul         1516
Jos Buttler       1509
ShikharDhawan    1392
ViratKohli       1385
SanjuSamson      1304
SuryakumarYadav  1225
GlennMaxwell     1214
Name: runs, dtype: int64
```

4. Top 10 batsmen based on past 3 years batting average.

In [239...]

```
# Total innings dismissed
innings_dismissed = df_batting_summary[df_batting_summary['out/not_out'] == 'out'].

# Total runs
total_runs = df_batting_summary['runs'].sum()

# Batting Average
batting_average = total_runs / innings_dismissed

print("Total Innings Dismissed:", innings_dismissed)
print("Total Runs Scored:", total_runs)
print("Batting Average:", batting_average)

import numpy as np
```

```
# Batting average calculation
```

```
batsman_batting_avg = df_batting_summary.groupby('batsmanName').apply(lambda x: np.
round(batsman_batting_avg,2).sort_values(ascending=False).head(10))
```

```
Total Innings Dismissed: 2527
Total Runs Scored: 64772
Batting Average: 25.63197467352592
batsmanName
VivrantSharma      69.00
SaurabhTiwary      57.50
KL Rahul           50.53
CameronGreen       50.22
RinkuSingh          49.85
HeinrichKlaasen     49.78
ChrisLynn           49.00
DevonConway         48.63
SaiSudharsan        46.09
FafduPlessis        43.60
dtype: float64
```

Out[239]:

5. Top 10 batsmen based on past 3 years strike rate.

In [240...

```
df_batting_summary.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3268 entries, 0 to 3267
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   match_id        3268 non-null   object
1   match           3268 non-null   object
2   teamInnings     3268 non-null   object
3   battingPos      3268 non-null   int64
4   batsmanName     3268 non-null   object
5   out/not_out     3268 non-null   object
6   runs            3268 non-null   int64
7   balls           3268 non-null   int64
8   4s              3268 non-null   int64
9   6s              3268 non-null   int64
10  SR              3268 non-null   object
dtypes: int64(5), object(6)
memory usage: 281.0+ KB
```

In [241...

```
# Leading batting average batsman
```

```
batting_average=df_batting_summary.groupby('batsmanName').apply(lambda x: np.nan i
round(batting_average,2).sort_values(ascending=False).head(10))
```

Out[241]:

```
batsmanName
MarkWood      220.00
PatCummins    195.00
DavidWiese    190.91
TimDavid      177.87
HeinrichKlaasen 177.08
RashidKhan    176.74
DhruvJurel    172.73
KrishnappaGowtham 171.43
MohammadNabi  170.00
AkashDeep     170.00
dtype: float64
```

6. Top 10 bowlers based on past 3 years total wickets taken.

In [242...]

Leading wicket takers

df_bowling_summary.head(5)

Out[242]:

	match_id	match	bowlingTeam	bowlerName	overs	maiden	runs	wickets	economy	0s
0	T203817	Super Kings Vs KKR	KKR	ShakibAlHasan	3.0	0	33	0	11.0	6
1	T203817	Super Kings Vs KKR	KKR	ShivamMavi	4.0	0	32	1	8.0	8
2	T203817	Super Kings Vs KKR	KKR	LockieFerguson	4.0	0	56	0	14.0	4
3	T203817	Super Kings Vs KKR	KKR	VarunChakravarthy	4.0	0	38	0	9.5	5
4	T203817	Super Kings Vs KKR	KKR	SunilNarine	4.0	0	26	2	6.5	6

In [243...]

```
wicket_taker=df_bowling_summary.groupby('bowlerName')['wickets'].sum()
wicket_taker.sort_values(ascending=False).head(10)
```

Out[243]:

```
bowlerName
MohammedShami      67
YuzvendraChahal    66
HarshalPatel       65
RashidKhan         63
AveshKhan          47
KagisoRabada       45
ArshdeepSingh     45
VarunChakravarthy  44
ShardulThakur     43
TrentBoult        42
Name: wickets, dtype: int64
```

7.Top 10 bowlers based on past 3 years bowling average.

In [244...]

```
bowling_average=df_bowling_summary.groupby('bowlerName').apply(lambda x: np.nan if
round(bowling_average,2).sort_values().head(10))
```

```
Out[244]: bowlerName
AyushBadoni      5.50
ImranTahir       8.00
DewaldBrevis     8.00
RamandeepSingh   9.00
MoisesHenriques  11.25
MarkWood         11.82
MohitSharma      13.37
TristanStubbs    14.00
ReeceTopley      14.00
MujeebUrRahman   14.50
dtype: float64
```

8.Top 10 bowlers based on past 3 years economy rate

```
In [245... economy_rate=df_bowling_summary.groupby('bowlerName')['economy'].mean()
round(economy_rate,2).sort_values().head(10)
```

```
Out[245]: bowlerName
AyushBadoni      3.67
ImranTahir       4.00
MoisesHenriques  5.60
PraveenDubey     6.33
DavidWilley      6.61
TilakVarma       6.67
MatthewShort     6.67
MitchellSantner  6.70
SunilNarine      6.72
TristanStubbs    7.00
Name: economy, dtype: float64
```

9.Top 5 batsmen based on past 3 years boundary runs (fours and sixes).

```
In [246... df_batting_summary['boundary_runs']=df_batting_summary['4s']*4+df_batting_summary['6s']*6
df_batting_summary.head(5)
```

```
Out[246]:
```

	match_id	match	teamInnings	battingPos	batsmanName	out/not_out	runs	balls	4s	6s
0	T203817	Super Kings Vs KKR	Super Kings	1	RuturajGaikwad	out	32	27	3	1
1	T203817	Super Kings Vs KKR	Super Kings	2	FafduPlessis	out	86	59	7	3
2	T203817	Super Kings Vs KKR	Super Kings	3	RobinUthappa	out	31	15	0	3
3	T203817	Super Kings Vs KKR	Super Kings	4	MoeenAli	not_out	37	20	2	3
4	T203817	Super Kings Vs KKR	KKR	1	ShubmanGill	out	51	43	6	0

```
In [247... boundary=df_batting_summary.groupby('batsmanName')['boundary_runs'].sum()
boundary.sort_values(ascending=False).head(10)
```

```
Out[247]: batsmanName
FafduPlessis      1108
ShubmanGill       1080
JosButtler        1040
RuturajGaikwad    974
KL Rahul          868
ShikharDhawan     864
SanjuSamson       854
YashasviJaiswal   844
SuryakumarYadav   836
GlennMaxwell      834
Name: boundary_runs, dtype: int64
```

10.Top 5 bowlers based on past 3 years dot ball %.

```
In [248... df_bowling_summary.head(5)
```

```
Out[248]:
```

	match_id	match	bowlingTeam	bowlerName	overs	maiden	runs	wickets	economy	0s
0	T203817	Super Kings Vs KKR	KKR	ShakibAlHasan	3.0	0	33	0	11.0	6
1	T203817	Super Kings Vs KKR	KKR	ShivamMavi	4.0	0	32	1	8.0	8
2	T203817	Super Kings Vs KKR	KKR	LockieFerguson	4.0	0	56	0	14.0	4
3	T203817	Super Kings Vs KKR	KKR	VarunChakravarthy	4.0	0	38	0	9.5	5
4	T203817	Super Kings Vs KKR	KKR	SunilNarine	4.0	0	26	2	6.5	6

```
In [249... dotball_pct=df_bowling_summary.groupby('bowlerName').apply(lambda x: np.nan if (x['round(dotball_pct,2)].sort_values(ascending=False).head(10)
```

```
Out[249]: bowlerName
ShreyasIyer      66.67
ReeceTopley      58.33
ImranTahir       58.33
DewaldBrevis     55.56
SimarjeetSingh   52.78
YudhvirSingh     52.08
MoisesHenriques  51.67
MohsinKhan       50.39
ChrisWoakes      50.00
MarkWood         50.00
dtype: float64
```

11.Top 4 teams based on past 3 years winning %.

In [250...

team_stats.head(4)

Out[250]:

	total_matches	total_matches_won	total_matches_loss	win%
Titans	33	23	10	69.70
Super Giants	29	17	12	58.62
Super Kings	45	25	20	55.56
RCB	45	25	20	55.56

12. Players year over year growth

In [254...

df_batting_summary

Out[254]:

	match_id	match	teamInnings	battingPos	batsmanName	out/not_out	runs	balls	4s	6s
0	T203817	Super Kings Vs KKR	Super Kings	1	RuturajGaikwad	out	32	27	3	1
1	T203817	Super Kings Vs KKR	Super Kings	2	FafduPlessis	out	86	59	7	3
2	T203817	Super Kings Vs KKR	Super Kings	3	RobinUthappa	out	31	15	0	3
3	T203817	Super Kings Vs KKR	Super Kings	4	MoeenAli	not_out	37	20	2	3
4	T203817	Super Kings Vs KKR	KKR	1	ShubmanGill	out	51	43	6	0
...
3263	T201566	Super Kings Vs Titans	Titans	3	SaiSudharsan	out	22	17	3	0
3264	T201566	Super Kings Vs Titans	Titans	4	HardikPandya	out	8	11	0	0
3265	T201566	Super Kings Vs Titans	Titans	5	VijayShankar	out	27	21	2	1
3266	T201566	Super Kings Vs Titans	Titans	6	RahulTewatia	not_out	15	14	1	1
3267	T201566	Super Kings Vs Titans	Titans	7	RashidKhan	not_out	10	3	1	1

3268 rows × 12 columns

In [276...

```
import pandas as pd

df_merged = df_batting_summary.merge(df_match_summary[['match_id', 'Year']], on='ma
df_merged.drop_duplicates('match_id',inplace=True)
player_runs_by_year = df_merged.groupby(['Year', 'batsmanName'])['runs'].sum().rese
```

In [277...

```
pd.set_option('display.max_rows', None)
print(player_runs_by_year)
```

```
# Display the entire DataFrame  
  
# Reset display options to default values  
pd.reset_option('display.max_rows')
```


	Year	batsmanName	runs
0	2021	DavidWarner	57
1	2021	DevduttPadikkal	21
2	2021	EvinLewis	124
3	2021	JasonRoy	56
4	2021	JosButtler	173
5	2021	KLRahul	313
6	2021	NitishRana	117
7	2021	PrabhsimranSingh	12
8	2021	PrithviShaw	138
9	2021	QuintondeKock	43
10	2021	RohitSharma	164
11	2021	RuturajGaikwad	362
12	2021	ShubmanGill	65
13	2021	StevenSmith	39
14	2021	VenkateshIyer	67
15	2021	ViratKohli	184
16	2022	AaronFinch	3
17	2022	AbhishekSharma	74
18	2022	AjinkyaRahane	9
19	2022	AnujRawat	4
20	2022	DavidWarner	0
21	2022	FafduPlessis	101
22	2022	IshanKishan	71
23	2022	JonnyBairstow	122
24	2022	JosButtler	534
25	2022	KLRahul	171
26	2022	MandeepSingh	0
27	2022	MatthewWade	32
28	2022	MayankAgarwal	104
29	2022	PrithviShaw	136
30	2022	PriyamGarg	4
31	2022	QuintondeKock	269
32	2022	RobinUthappa	65
33	2022	RohitSharma	44
34	2022	RuturajGaikwad	292
35	2022	ShikharDhawan	8
36	2022	VenkateshIyer	56
37	2022	ViratKohli	120
38	2022	WriddhimanSaha	82
39	2022	YashasviJaiswal	85
40	2023	AbhishekSharma	78
41	2023	AnmolpreetSingh	67
42	2023	CameronGreen	6
43	2023	DavidWarner	155
44	2023	DeepakHooda	5
45	2023	DevonConway	1
46	2023	HarryBrook	118
47	2023	IshanKishan	46
48	2023	JasonRoy	129
49	2023	KLRahul	133
50	2023	KaranSharma	3
51	2023	MatthewShort	11
52	2023	NarayanJagadeesan	19
53	2023	PhilSalt	0
54	2023	PrabhsimranSingh	209
55	2023	RahmanullahGurbaz	65
56	2023	RohitSharma	50
57	2023	RuturajGaikwad	312
58	2023	ViratKohli	376
59	2023	VivrantSharma	69
60	2023	WriddhimanSaha	225
61	2023	YashasviJaiswal	374