

In [1]: `import os`

In [2]: `os.listdir()`

Out[2]: `['.ipynb_checkpoints',  
'IPL Ball-by-Ball 2008-2020.csv',  
'IPL Matches 2008-2020.csv',  
'Untitled.ipynb']`

In [4]: `os.listdir('IPL')`

```
-----
FileNotFoundError                                Traceback (most recent call last)
<ipython-input-4-10c6061b1eab> in <module>
----> 1 os.listdir('IPL')

FileNotFoundError: [WinError 3] The system cannot find the path specified: 'IPL'
```

In [5]: `import pandas as pd  
import numpy as np  
import seaborn as sns  
import matplotlib.pyplot as plt`

In [6]: `import datetime`

In [7]: `match_df = pd.read_csv('IPL Matches 2008-2020.csv')`

In [14]: `ball_df = pd.read_csv('IPL Ball-by-Ball 2008-2020.csv')`

In [20]: `match_df.head()`

Out[20]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2
0	335982	Bangalore	2008-04-18	BB McCullum	Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkata Knight Riders
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Chennai Super Kings
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	Rajasthan Royals
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Royal Challengers Bangalore
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	Deccan Chargers

In [21]:

ball\_df.head()

Out[21]:

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total_runs
0	335982	1	6	5	RT Ponting	BB McCullum	AA Noffke	1	0	1
1	335982	1	6	6	BB McCullum	RT Ponting	AA Noffke	1	0	1
2	335982	1	7	1	BB McCullum	RT Ponting	Z Khan	0	0	0
3	335982	1	7	2	BB McCullum	RT Ponting	Z Khan	1	0	1
4	335982	1	7	3	RT Ponting	BB McCullum	Z Khan	1	0	1

In [22]:

match\_df.tail()

Out[22]:

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	t
811	1216547	Dubai	2020-09-28	AB de Villiers	Dubai International Cricket Stadium	0	Royal Challengers Bangalore	Mumbai Indians	
812	1237177	Dubai	2020-11-05	JJ Bumrah	Dubai International Cricket Stadium	0	Mumbai Indians	Delhi Capitals	
813	1237178	Abu Dhabi	2020-11-06	KS Williamson	Sheikh Zayed Stadium	0	Royal Challengers Bangalore	Sunrisers Hyderabad	
814	1237180	Abu Dhabi	2020-11-08	MP Stoinis	Sheikh Zayed Stadium	0	Delhi Capitals	Sunrisers Hyderabad	
815	1237181	Dubai	2020-11-10	TA Boult	Dubai International Cricket Stadium	0	Delhi Capitals	Mumbai Indians	

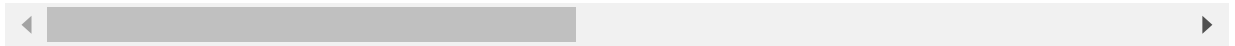
In [23]:

ball\_df.tail()

Out[23]:

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total
193463	1237181	1	12	5	RR Pant	SS Iyer	NM Coulter-Nile	0	0	

	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total_runs
<b>193464</b>	1237181	1	12	6	RR Pant	SS Iyer	NM Coulter-Nile	1	0	
<b>193465</b>	1237181	1	13	1	RR Pant	SS Iyer	KH Pandya	0	1	
<b>193466</b>	1237181	1	13	2	RR Pant	SS Iyer	KH Pandya	1	0	
<b>193467</b>	1237181	1	13	3	SS Iyer	RR Pant	KH Pandya	1	0	



In [24]:

```
match_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 816 entries, 0 to 815
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                     816 non-null   int64
1   city                   803 non-null   object
2   date                   816 non-null   object
3   player_of_match       812 non-null   object
4   venue                  816 non-null   object
5   neutral_venue         816 non-null   int64
6   team1                  816 non-null   object
7   team2                  816 non-null   object
8   toss_winner            816 non-null   object
9   toss_decision         816 non-null   object
10  winner                 812 non-null   object
11  result                 812 non-null   object
12  result_margin         799 non-null   float64
13  eliminator            812 non-null   object
14  method                 19 non-null    object
15  umpire1                816 non-null   object
16  umpire2                816 non-null   object
dtypes: float64(1), int64(2), object(14)
memory usage: 108.5+ KB
```

In [26]:

```
match_df.columns
```

```
Out[26]: Index(['id', 'city', 'date', 'player_of_match', 'venue', 'neutral_venue',
               'team1', 'team2', 'toss_winner', 'toss_decision', 'winner', 'result',
               'result_margin', 'eliminator', 'method', 'umpire1', 'umpire2'],
              dtype='object')
```

In [28]:

```
match_df.items()
```

```
Out[28]: <generator object DataFrame.items at 0x000001D6E6D5E580>
```

In [30]:

```
match_df.shape
```

```
Out[30]: (816, 17)
```

In [36]:

```
ball_df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 193468 entries, 0 to 193467
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   id                    193468 non-null  int64
 1   inning                193468 non-null  int64
 2   over                  193468 non-null  int64
 3   ball                  193468 non-null  int64
 4   batsman               193468 non-null  object
 5   non_striker           193468 non-null  object
 6   bowler                193468 non-null  object
 7   batsman_runs          193468 non-null  int64
 8   extra_runs            193468 non-null  int64
 9   total_runs            193468 non-null  int64
10   non_boundary          193468 non-null  int64
11   is_wicket             193468 non-null  int64
12   dismissal_kind        9495 non-null    object
13   player_dismissed     9495 non-null    object
14   fielder               6784 non-null    object
15   extras_type           10233 non-null   object
16   batting_team          193468 non-null  object
17   bowling_team          193277 non-null  object
dtypes: int64(9), object(9)
memory usage: 26.6+ MB

```

```
In [38]: ball_df.shape
```

```
Out[38]: (193468, 18)
```

```
In [41]: ball_df.columns
```

```
Out[41]: Index(['id', 'inning', 'over', 'ball', 'batsman', 'non_striker', 'bowler',
               'batsman_runs', 'extra_runs', 'total_runs', 'non_boundary', 'is_wicket',
               'dismissal_kind', 'player_dismissed', 'fielder', 'extras_type',
               'batting_team', 'bowling_team'],
              dtype='object')
```

```
In [ ]:
```