

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression

match=pd.read_csv(r"C:\Users\Admin\Downloads\matches.csv")
match.head()
```

	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
1	Rising Pune Supergiant	Rising Pune Supergiant
2	Kolkata Knight Riders	Kolkata Knight Riders
3	Kings XI Punjab	Kings XI Punjab
4	Delhi Daredevils	Royal Challengers Bangalore

	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

```
match.replace('supergiants','supergiant ',inplace=True)
```

```
match.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 756 entries, 0 to 755
```

```
Data columns (total 18 columns):
```

#	Column	Non-Null Count	Dtype
0	id	756 non-null	int64
1	season	756 non-null	int64
2	city	749 non-null	object
3	date	756 non-null	object
4	team1	756 non-null	object
5	team2	756 non-null	object
6	toss_winner	756 non-null	object
7	toss_decision	756 non-null	object
8	result	756 non-null	object
9	dl_applied	756 non-null	int64
10	winner	752 non-null	object
11	win_by_runs	756 non-null	int64
12	win_by_wickets	756 non-null	int64
13	player_of_match	752 non-null	object
14	venue	756 non-null	object
15	umpire1	754 non-null	object
16	umpire2	754 non-null	object
17	umpire3	119 non-null	object

```
dtypes: int64(5), object(13)
```

```
memory usage: 106.4+ KB
```

```
match.describe()
```

	id	season	dl_applied	win_by_runs
win_by_wickets				
count	756.000000	756.000000	756.000000	756.000000
mean	1792.178571	2013.444444	0.025132	13.283069
std	3464.478148	3.366895	0.156630	23.471144
min	1.000000	2008.000000	0.000000	0.000000

25%	189.750000	2011.000000	0.000000	0.000000
0.000000				
50%	378.500000	2013.000000	0.000000	0.000000
4.000000				
75%	567.250000	2016.000000	0.000000	19.000000
6.000000				
max	11415.000000	2019.000000	1.000000	146.000000
10.000000				

match.dtypes

id	int64
season	int64
city	object
date	object
team1	object
team2	object
toss_winner	object
toss_decision	object
result	object
dl_applied	int64
winner	object
win_by_runs	int64
win_by_wickets	int64
player_of_match	object
venue	object
umpire1	object
umpire2	object
umpire3	object
dtype:	object

match.shape

(756, 18)

match.isnull().sum()

id	0
season	0
city	7
date	0
team1	0
team2	0
toss_winner	0
toss_decision	0
result	0
dl_applied	0
winner	4
win_by_runs	0
win_by_wickets	0
player_of_match	4

```
venue          0
umpire1        2
umpire2        2
umpire3        2
umpire3        637
dtype: int64
```

```
match.dropna(0)
```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\2959912194.py:1:
FutureWarning: In a future version of pandas all arguments of
DataFrame.dropna will be keyword-only.
  match.dropna(0)
```

	id	season	city	date	
team1 \					
636	7894	2018	Mumbai	07/04/18	Mumbai Indians
637	7895	2018	Mohali	08/04/18	Delhi Daredevils
638	7896	2018	Kolkata	08/04/18	Royal Challengers Bangalore
639	7897	2018	Hyderabad	09/04/18	Rajasthan Royals
640	7898	2018	Chennai	10/04/18	Kolkata Knight Riders
..
..					
750	11346	2019	Mohali	05/05/19	Chennai Super Kings
751	11347	2019	Mumbai	05/05/19	Kolkata Knight Riders
752	11412	2019	Chennai	07/05/19	Chennai Super Kings
754	11414	2019	Visakhapatnam	10/05/19	Delhi Capitals
755	11415	2019	Hyderabad	12/05/19	Mumbai Indians

	team2	toss_winner	toss_decision
result \			
636	Chennai Super Kings	Chennai Super Kings	field normal
637	Kings XI Punjab	Kings XI Punjab	field normal
638	Kolkata Knight Riders	Kolkata Knight Riders	field normal
639	Sunrisers Hyderabad	Sunrisers Hyderabad	field normal
640	Chennai Super Kings	Chennai Super Kings	field normal

```

..          ...          ...          ...          ..
.
750      Kings XI Punjab      Kings XI Punjab      field
normal
751      Mumbai Indians      Mumbai Indians      field
normal
752      Mumbai Indians      Chennai Super Kings      bat
normal
754      Chennai Super Kings      Chennai Super Kings      field
normal
755      Chennai Super Kings      Mumbai Indians      bat
normal

```

```

      dl_applied      winner      win_by_runs      win_by_wickets \
636      0      Chennai Super Kings      0      1
637      0      Kings XI Punjab      0      6
638      0      Kolkata Knight Riders      0      4
639      0      Sunrisers Hyderabad      0      9
640      0      Chennai Super Kings      0      5
..          ...          ...          ...          ...
750      0      Kings XI Punjab      0      6
751      0      Mumbai Indians      0      9
752      0      Mumbai Indians      0      6
754      0      Chennai Super Kings      0      6
755      0      Mumbai Indians      1      0

```

```

      player_of_match      venue
\
636      DJ Bravo      Wankhede Stadium
637      KL Rahul      Punjab Cricket Association IS Bindra Stadium, ...
638      SP Narine      Eden Gardens
639      S Dhawan      Rajiv Gandhi International Stadium, Uppal
640      SW Billings      MA Chidambaram Stadium, Chepauk
..          ...          ...
750      KL Rahul      IS Bindra Stadium
751      HH Pandya      Wankhede Stadium
752      AS Yadav      M. A. Chidambaram Stadium
754      F du Plessis      ACA-VDCA Stadium
755      JJ Bumrah      Rajiv Gandhi Intl. Cricket Stadium

```

	umpire1	umpire2	umpire3
636	Chris Gaffaney	A Nanda Kishore	Anil Chaudhary
637	Rod Tucker	K Ananthapadmanabhan	Nitin Menon
638	C Shamshuddin	A.D Deshmukh	S Ravi
639	Nigel Llong	Vineet Kulkarni	0 Nandan
640	Anil Chaudhary	Chris Gaffaney	A Nanda Kishore
..
750	KN Ananthapadmanabhan	C Shamshuddin	Bruce Oxenford
751	Nanda Kishore	0 Nandan	S Ravi
752	Nigel Llong	Nitin Menon	Ian Gould
754	Sundaram Ravi	Bruce Oxenford	Chettithody Shamshuddin
755	Nitin Menon	Ian Gould	Nigel Llong

[118 rows x 18 columns]

```

match['winner'].replace('',np.nan, inplace=True)
match['city'].replace('',np.nan, inplace=True)
match['player_of_match'].replace('',np.nan, inplace=True)
match['umpire1'].replace('',np.nan, inplace=True)
match['umpire2'].replace('',np.nan, inplace=True)
match['umpire3'].replace('',np.nan, inplace=True)
match.dropna(subset=['winner'],inplace=True)
match.dropna(subset=['city'],inplace=True)
match.dropna(subset=['player_of_match'],inplace=True)
match.dropna(subset=['umpire1'],inplace=True)
match.dropna(subset=['umpire2'],inplace=True)
match.dropna(subset=['umpire3'],inplace=True)

```

```
match.isnull().sum()
```

```
id          0
season      0
city        0
date        0
team1       0
team2       0
toss_winner 0
toss_decision 0
result      0
dl_applied  0
winner      0
win_by_runs 0
win_by_wickets 0
player_of_match 0
venue       0
umpire1     0
umpire2     0
umpire3     0
dtype: int64
```

```
delivery=pd.read_csv(r"C:\Users\Admin\Downloads\deliveries.csv\
deliveries.csv")
delivery.head()
```

	match_id	inning	batting_team		bowling_team		
over \							
0	1	1	Sunrisers	Hyderabad	Royal Challengers	Bangalore	
1							
1	1	1	Sunrisers	Hyderabad	Royal Challengers	Bangalore	
1							
2	1	1	Sunrisers	Hyderabad	Royal Challengers	Bangalore	
1							
3	1	1	Sunrisers	Hyderabad	Royal Challengers	Bangalore	
1							
4	1	1	Sunrisers	Hyderabad	Royal Challengers	Bangalore	
1							

	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs
\							
0	1	DA Warner	S Dhawan	TS Mills	0	...	0
1	2	DA Warner	S Dhawan	TS Mills	0	...	0
2	3	DA Warner	S Dhawan	TS Mills	0	...	0
3	4	DA Warner	S Dhawan	TS Mills	0	...	0
4	5	DA Warner	S Dhawan	TS Mills	0	...	0

	legbye_runs	noball_runs	penalty_runs	batsman_runs	extra_runs	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	4	0	
3	0	0	0	0	0	
4	0	0	0	0	2	

	total_runs	player_dismissed	dismissal_kind	fielder
0	0	NaN	NaN	NaN
1	0	NaN	NaN	NaN
2	4	NaN	NaN	NaN
3	0	NaN	NaN	NaN
4	2	NaN	NaN	NaN

[5 rows x 21 columns]

delivery.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 179078 entries, 0 to 179077
Data columns (total 21 columns):
```

#	Column	Non-Null	Count	Dtype
0	match_id	179078	non-null	int64
1	inning	179078	non-null	int64
2	batting_team	179078	non-null	object
3	bowling_team	179078	non-null	object
4	over	179078	non-null	int64
5	ball	179078	non-null	int64
6	batsman	179078	non-null	object
7	non_striker	179078	non-null	object
8	bowler	179078	non-null	object
9	is_super_over	179078	non-null	int64
10	wide_runs	179078	non-null	int64
11	bye_runs	179078	non-null	int64
12	legbye_runs	179078	non-null	int64
13	noball_runs	179078	non-null	int64
14	penalty_runs	179078	non-null	int64
15	batsman_runs	179078	non-null	int64
16	extra_runs	179078	non-null	int64
17	total_runs	179078	non-null	int64
18	player_dismissed	8834	non-null	object
19	dismissal_kind	8834	non-null	object
20	fielder	6448	non-null	object

dtypes: int64(13), object(8)

memory usage: 28.7+ MB

delivery.replace('supergiants','supergiant ',inplace=True)


```
delivery.shape
```

```
(179078, 21)
```

```
delivery.dtypes
```

```
match_id      int64
inning        int64
batting_team   object
bowling_team   object
over          int64
ball          int64
batsman        object
non_striker    object
bowler         object
is_super_over  int64
wide_runs      int64
bye_runs       int64
legbye_runs    int64
noball_runs    int64
penalty_runs   int64
batsman_runs   int64
extra_runs     int64
total_runs     int64
player_dismissed object
dismissal_kind object
fielder        object
dtype: object
```

```
delivery.isnull().sum()
```

```
match_id      0
inning        0
batting_team   0
bowling_team   0
over          0
ball          0
batsman        0
non_striker    0
bowler         0
is_super_over  0
wide_runs      0
bye_runs       0
legbye_runs    0
noball_runs    0
penalty_runs   0
batsman_runs   0
extra_runs     0
total_runs     0
player_dismissed 170244
```

```
dismissal_kind      170244
fielder              172630
dtype: int64
```

```
delivery.dropna(0)
```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\2485708266.py:1:
FutureWarning: In a future version of pandas all arguments of
DataFrame.dropna will be keyword-only.
  delivery.dropna(0)
```

	match_id	inning	batting_team \
11	1	1	Sunrisers Hyderabad
64	1	1	Sunrisers Hyderabad
94	1	1	Sunrisers Hyderabad
165	1	2	Royal Challengers Bangalore
199	1	2	Royal Challengers Bangalore
...
178977	11415	2	Chennai Super Kings
179019	11415	2	Chennai Super Kings
179033	11415	2	Chennai Super Kings
179067	11415	2	Chennai Super Kings
179075	11415	2	Chennai Super Kings

	bowling_team	over	ball	batsman
non_striker \				
11	Royal Challengers Bangalore	2	5	DA Warner S
Dhawan				
64	Royal Challengers Bangalore	11	3	S Dhawan MC
Henriques				
94	Royal Challengers Bangalore	16	2	MC Henriques Yuvraj
Singh				
165	Sunrisers Hyderabad	7	3	CH Gayle TM
Head				
199	Sunrisers Hyderabad	12	4	KM Jadhav TM
Head				
...
...				
178977	Mumbai Indians	4	6	F du Plessis SR
Watson				
179019	Mumbai Indians	11	3	AT Rayudu SR
Watson				
179033	Mumbai Indians	13	5	SR Watson MS
Dhoni				
179067	Mumbai Indians	19	2	DJ Bravo SR
Watson				
179075	Mumbai Indians	20	4	SR Watson RA
Jadeja				

```
bowler is_super_over ... bye_runs legbye_runs
```

noball_runs	\				
11	A Choudhary	0	...	0	0
0					
64	STR Binny	0	...	0	0
0					
94	YS Chahal	0	...	0	0
0					
165	DJ Hooda	0	...	0	0
0					
199	MC Henriques	0	...	0	0
0					
...
...					
178977	KH Pandya	0	...	0	0
0					
179019	JJ Bumrah	0	...	0	0
0					
179033	HH Pandya	0	...	0	0
0					
179067	JJ Bumrah	0	...	0	0
0					
179075	SL Malinga	0	...	0	0
0					

	penalty_runs	batsman_runs	extra_runs	total_runs	
player_dismissed	\				
11	0	0	0	0	DA Warner
64	0	0	0	0	S Dhawan
94	0	0	0	0	MC Henriques
165	0	0	0	0	CH Gayle
199	0	1	0	1	KM Jadhav
...	
...					
178977	0	0	0	0	F du Plessis
179019	0	0	0	0	AT Rayudu
179033	0	1	0	1	MS Dhoni
179067	0	0	0	0	DJ Bravo
179075	0	1	0	1	SR Watson

	dismissal_kind	fielder
11	caught	Mandeep Singh
64	caught	Sachin Baby
94	caught	Sachin Baby
165	caught	DA Warner
199	run out	BCJ Cutting
...
178977	stumped	Q de Kock
179019	caught	Q de Kock
179033	run out	Ishan Kishan
179067	caught	Q de Kock
179075	run out	KH Pandya

[6448 rows x 21 columns]

```
total_score_df=delivery.groupby(['match_id','inning']).sum()
['total_runs']
total_score_df
```

match_id	inning	
1	1	207
	2	172
2	1	184
	2	187
3	1	183
		...
11413	2	170
11414	1	155
	2	162
11415	1	152
	2	157

Name: total_runs, Length: 1528, dtype: int64

```
total_score_df=delivery.groupby(['match_id','inning']).sum()
['total_runs'].reset_index()
total_score_df
```

	match_id	inning	total_runs
0	1	1	207
1	1	2	172
2	2	1	184
3	2	2	187
4	3	1	183
...
1523	11413	2	170
1524	11414	1	155
1525	11414	2	162
1526	11415	1	152
1527	11415	2	157

```
[1528 rows x 3 columns]
```

```
total_score_df=total_score_df[total_score_df['inning']==1]
total_score_df
```

	match_id	inning	total_runs
0	1	1	207
2	2	1	184
4	3	1	183
6	4	1	163
8	5	1	157
...
1518	11347	1	143
1520	11412	1	136
1522	11413	1	171
1524	11414	1	155
1526	11415	1	152

```
[756 rows x 3 columns]
```

```
match_df=match.merge(total_score_df[['match_id','total_runs']],left_on=
='id',right_on='match_id')
match_df
```

	id	season	city	date	
team1 \					
0	7894	2018	Mumbai	07/04/18	Mumbai Indians
1	7895	2018	Mohali	08/04/18	Delhi Daredevils
2	7896	2018	Kolkata	08/04/18	Royal Challengers Bangalore
3	7897	2018	Hyderabad	09/04/18	Rajasthan Royals
4	7898	2018	Chennai	10/04/18	Kolkata Knight Riders
..
..					
113	11346	2019	Mohali	05/05/19	Chennai Super Kings
114	11347	2019	Mumbai	05/05/19	Kolkata Knight Riders
115	11412	2019	Chennai	07/05/19	Chennai Super Kings
116	11414	2019	Visakhapatnam	10/05/19	Delhi Capitals
117	11415	2019	Hyderabad	12/05/19	Mumbai Indians

result \	team2	toss_winner	toss_decision
0	Chennai Super Kings	Chennai Super Kings	field
normal			
1	Kings XI Punjab	Kings XI Punjab	field
normal			
2	Kolkata Knight Riders	Kolkata Knight Riders	field
normal			
3	Sunrisers Hyderabad	Sunrisers Hyderabad	field
normal			
4	Chennai Super Kings	Chennai Super Kings	field
normal			
..
.			
113	Kings XI Punjab	Kings XI Punjab	field
normal			
114	Mumbai Indians	Mumbai Indians	field
normal			
115	Mumbai Indians	Chennai Super Kings	bat
normal			
116	Chennai Super Kings	Chennai Super Kings	field
normal			
117	Chennai Super Kings	Mumbai Indians	bat
normal			
dl_applied	winner	win_by_runs	win_by_wickets \
0	Chennai Super Kings	0	1
1	Kings XI Punjab	0	6
2	Kolkata Knight Riders	0	4
3	Sunrisers Hyderabad	0	9
4	Chennai Super Kings	0	5
..
113	Kings XI Punjab	0	6
114	Mumbai Indians	0	9
115	Mumbai Indians	0	6
116	Chennai Super Kings	0	6
117	Mumbai Indians	1	0
player_of_match	venue		
\			
0	DJ Bravo	Wankhede Stadium	
1	KL Rahul	Punjab Cricket Association IS Bindra Stadium, ...	
2	SP Narine	Eden Gardens	
3	S Dhawan	Rajiv Gandhi International Stadium, Uppal	
4	SW Billings	MA Chidambaram Stadium, Chepauk	

..
113	KL Rahul	IS Bindra Stadium
114	HH Pandya	Wankhede Stadium
115	AS Yadav	M. A. Chidambaram Stadium
116	F du Plessis	ACA-VDCA Stadium
117	JJ Bumrah	Rajiv Gandhi Intl. Cricket Stadium
umpire1 umpire2		
umpire3 \		
0	Chris Gaffaney	A Nanda Kishore Anil
Chaudhary		
1	Rod Tucker	K Ananthapadmanabhan Nitin
Menon		
2	C Shamshuddin	A.D Deshmukh S
Ravi		
3	Nigel Llong	Vineet Kulkarni 0
Nandan		
4	Anil Chaudhary	Chris Gaffaney A Nanda
Kishore		
..
...		
113	KN Ananthapadmanabhan	C Shamshuddin Bruce
Oxenford		
114	Nanda Kishore	0 Nandan S
Ravi		
115	Nigel Llong	Nitin Menon Ian
Gould		
116	Sundaram Ravi	Bruce Oxenford Chettithody
Shamshuddin		
117	Nitin Menon	Ian Gould Nigel
Llong		
match_id total_runs		
0	7894	169
1	7895	172
2	7896	182
3	7897	133
4	7898	207
..
113	11346	172
114	11347	143
115	11412	136
116	11414	155
117	11415	152

```
[118 rows x 20 columns]
```

```
#to find how many teams
```

```
match_df['team1'].unique()
```

```
array(['Mumbai Indians', 'Delhi Daredevils',  
      'Royal Challengers Bangalore', 'Rajasthan Royals',  
      'Kolkata Knight Riders', 'Kings XI Punjab', 'Chennai Super  
Kings',  
      'Sunrisers Hyderabad', 'Delhi Capitals'], dtype=object)
```

```
team=['Sunrisers Hyderabad', 'Mumbai Indians', 'Royal Challengers  
Bangalore', 'Kolkata Knight Riders', 'Kings XI Punjab',  
      'Chennai Super Kings', 'Rajasthan Royals', 'Delhi  
Capitals', 'Gujarat Titans', 'Lucknow Super Giants']
```

```
match_df['team1']=match_df['team1'].str.replace('Delhi  
Daredevils', 'Delhi Capitals')  
match_df['team2']=match_df['team2'].str.replace('Delhi  
Daredevils', 'Delhi Capitals')
```

```
match_df['team1']=match_df['team1'].str.replace('Deccan  
Chargers', 'Sunrisers Hyderabad')  
match_df['team2']=match_df['team2'].str.replace('Deccan  
Chargers', 'Sunrisers Hyderabad')
```

```
match_df['team1']=match_df['team1'].str.replace('Pune  
Warriors', 'Lucknow Super Giants')  
match_df['team2']=match_df['team2'].str.replace('Pune  
Warriors', 'Lucknow Super Giants')
```

```
match_df['team1']=match_df['team1'].str.replace('Gujarat  
Lions', 'Gujarat Titans')  
match_df['team2']=match_df['team2'].str.replace('Gujarat  
Lions', 'Gujarat Titans')
```

```
match_df=match_df[match_df['team1'].isin(team)]  
match_df=match_df[match_df['team2'].isin(team)]
```

```
match_df=match_df[match_df['team1'].isin(team)]  
match_df=match_df[match_df['team2'].isin(team)]
```


#duckworth applied method

match_df=match_df[match_df['dl_applied']==0]

match_df

	id	season	city	date	
team1 \					
0	7894	2018	Mumbai	07/04/18	Mumbai
Indians					
1	7895	2018	Mohali	08/04/18	Delhi
Capitals					
2	7896	2018	Kolkata	08/04/18	Royal Challengers
Bangalore					
3	7897	2018	Hyderabad	09/04/18	Rajasthan
Royals					
4	7898	2018	Chennai	10/04/18	Kolkata Knight
Riders					
..
..					
113	11346	2019	Mohali	05/05/19	Chennai Super
Kings					
114	11347	2019	Mumbai	05/05/19	Kolkata Knight
Riders					
115	11412	2019	Chennai	07/05/19	Chennai Super
Kings					
116	11414	2019	Visakhapatnam	10/05/19	Delhi
Capitals					
117	11415	2019	Hyderabad	12/05/19	Mumbai
Indians					
			team2	toss_winner	toss_decision
result \					
0		Chennai Super Kings	Chennai Super Kings		field
normal					
1		Kings XI Punjab	Kings XI Punjab		field
normal					
2		Kolkata Knight Riders	Kolkata Knight Riders		field
normal					
3		Sunrisers Hyderabad	Sunrisers Hyderabad		field
normal					
4		Chennai Super Kings	Chennai Super Kings		field
normal					
..	
.					..
113		Kings XI Punjab	Kings XI Punjab		field
normal					
114		Mumbai Indians	Mumbai Indians		field
normal					
115		Mumbai Indians	Chennai Super Kings		bat
normal					
116		Chennai Super Kings	Chennai Super Kings		field

normal					
117	Chennai Super Kings	Mumbai Indians		bat	
normal					
	dl_applied	winner	win_by_runs	win_by_wickets	\
0	0	Chennai Super Kings	0	1	
1	0	Kings XI Punjab	0	6	
2	0	Kolkata Knight Riders	0	4	
3	0	Sunrisers Hyderabad	0	9	
4	0	Chennai Super Kings	0	5	
..	
113	0	Kings XI Punjab	0	6	
114	0	Mumbai Indians	0	9	
115	0	Mumbai Indians	0	6	
116	0	Chennai Super Kings	0	6	
117	0	Mumbai Indians	1	0	
	player_of_match			venue	
\					
0	DJ Bravo			Wankhede Stadium	
1	KL Rahul	Punjab Cricket Association IS Bindra Stadium, ...			
2	SP Narine			Eden Gardens	
3	S Dhawan	Rajiv Gandhi International Stadium, Uppal			
4	SW Billings	MA Chidambaram Stadium, Chepauk			
..	
113	KL Rahul			IS Bindra Stadium	
114	HH Pandya			Wankhede Stadium	
115	AS Yadav	M. A. Chidambaram Stadium			
116	F du Plessis			ACA-VDCA Stadium	
117	JJ Bumrah	Rajiv Gandhi Intl. Cricket Stadium			
	umpire1	umpire2			
umpire3 \					
0	Chris Gaffaney	A Nanda Kishore		Anil	
Chaudhary					
1	Rod Tucker	K Ananthapadmanabhan		Nitin	
Menon					
2	C Shamshuddin	A.D Deshmukh		S	
Ravi					
3	Nigel Llong	Vineet Kulkarni		0	

Nandan			
4	Anil Chaudhary	Chris Gaffaney	A Nanda
Kishore			
..	
...			
113	KN Ananthapadmanabhan	C Shamshuddin	Bruce
Oxenford			
114	Nanda Kishore	O Nandan	S
Ravi			
115	Nigel Llong	Nitin Menon	Ian
Gould			
116	Sundaram Ravi	Bruce Oxenford	Chettithody
Shamshuddin			
117	Nitin Menon	Ian Gould	Nigel
Llong			

	match_id	total_runs
0	7894	169
1	7895	172
2	7896	182
3	7897	133
4	7898	207
..
113	11346	172
114	11347	143
115	11412	136
116	11414	155
117	11415	152

[115 rows x 20 columns]

```
match_df=match_df[['match_id','city','winner','total_runs']]
match_df
```

	match_id	city	winner	total_runs
0	7894	Mumbai	Chennai Super Kings	169
1	7895	Mohali	Kings XI Punjab	172
2	7896	Kolkata	Kolkata Knight Riders	182
3	7897	Hyderabad	Sunrisers Hyderabad	133
4	7898	Chennai	Chennai Super Kings	207
..
113	11346	Mohali	Kings XI Punjab	172
114	11347	Mumbai	Mumbai Indians	143
115	11412	Chennai	Mumbai Indians	136
116	11414	Visakhapatnam	Chennai Super Kings	155
117	11415	Hyderabad	Mumbai Indians	152

[115 rows x 4 columns]

```
delivery_df=match_df.merge(delivery,on='match_id')
delivery_df
```

	match_id	city	winner	total_runs_x	inning
\					
0	7894	Mumbai	Chennai Super Kings	169	1
1	7894	Mumbai	Chennai Super Kings	169	1
2	7894	Mumbai	Chennai Super Kings	169	1
3	7894	Mumbai	Chennai Super Kings	169	1
4	7894	Mumbai	Chennai Super Kings	169	1
...
27783	11415	Hyderabad	Mumbai Indians	152	2
27784	11415	Hyderabad	Mumbai Indians	152	2
27785	11415	Hyderabad	Mumbai Indians	152	2
27786	11415	Hyderabad	Mumbai Indians	152	2
27787	11415	Hyderabad	Mumbai Indians	152	2
...					
0					
...					
1					
...					
2					
...					
3					
...					
4					
...					
...					
...					
27783					
...					
27784					
...					
27785					
...					
27786					
...					
27787					

...

	bye_runs	legbye_runs	noball_runs	penalty_runs	batsman_runs	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	4
4	0	0	0	0	0	1
...
27783	0	0	0	0	0	1
27784	0	0	0	0	0	2
27785	0	0	0	0	0	1
27786	0	0	0	0	0	2
27787	0	0	0	0	0	0

	extra_runs	total_runs_y	player_dismissed	dismissal_kind	
fielder					
0	0	0	NaN	NaN	
NaN					
1	0	0	NaN	NaN	
NaN					
2	0	0	NaN	NaN	
NaN					
3	0	4	NaN	NaN	
NaN					
4	0	1	NaN	NaN	
NaN					
...
...					
27783	0	1	NaN	NaN	
NaN					
27784	0	2	NaN	NaN	
NaN					
27785	0	1	SR Watson	run out	KH
Pandya					
27786	0	2	NaN	NaN	
NaN					
27787	0	0	SN Thakur	lbw	
NaN					

[27788 rows x 24 columns]

```
delivery_df=delivery_df[delivery_df['inning']==2]
delivery_df
```

	match_id	city	winner	total_runs_x	inning
\					
122	7894	Mumbai	Chennai Super Kings	169	2
123	7894	Mumbai	Chennai Super Kings	169	2

124	7894	Mumbai	Chennai Super Kings	169	2
125	7894	Mumbai	Chennai Super Kings	169	2
126	7894	Mumbai	Chennai Super Kings	169	2
...
27783	11415	Hyderabad	Mumbai Indians	152	2
27784	11415	Hyderabad	Mumbai Indians	152	2
27785	11415	Hyderabad	Mumbai Indians	152	2
27786	11415	Hyderabad	Mumbai Indians	152	2
27787	11415	Hyderabad	Mumbai Indians	152	2
<div> <div></div> <div>batting_team</div> <div>bowling_team</div> <div>over</div> <div>ball</div> <div>batsman</div> <div>...</div> </div>					
122	Chennai Super Kings	Mumbai Indians	1	1	SR Watson ...
123	Chennai Super Kings	Mumbai Indians	1	2	SR Watson ...
124	Chennai Super Kings	Mumbai Indians	1	3	AT Rayudu ...
125	Chennai Super Kings	Mumbai Indians	1	4	SR Watson ...
126	Chennai Super Kings	Mumbai Indians	1	5	AT Rayudu ...
...
27783	Chennai Super Kings	Mumbai Indians	20	2	RA Jadeja ...
27784	Chennai Super Kings	Mumbai Indians	20	3	SR Watson ...
27785	Chennai Super Kings	Mumbai Indians	20	4	SR Watson ...
27786	Chennai Super Kings	Mumbai Indians	20	5	SN Thakur ...
27787	Chennai Super Kings	Mumbai Indians	20	6	SN Thakur ...
<div> <div>bye_runs</div> <div>legbye_runs</div> <div>noball_runs</div> <div>penalty_runs</div> <div>batsman_runs</div> <div>\</div> </div>					
122	0	0	0	0	0
123	0	0	0	0	1
124	0	0	0	0	1
125	0	0	0	0	1
126	0	0	0	0	0

```

...      ...      ...      ...      ...      ...
27783      0      0      0      0      1
27784      0      0      0      0      2
27785      0      0      0      0      1
27786      0      0      0      0      2
27787      0      0      0      0      0

```

```

      extra_runs  total_runs_y  player_dismissed  dismissal_kind
fielder
122      0      0      NaN      NaN
NaN
123      0      1      NaN      NaN
NaN
124      0      1      NaN      NaN
NaN
125      0      1      NaN      NaN
NaN
126      0      0      NaN      NaN
NaN
...      ...      ...      ...      ...
...
27783      0      1      NaN      NaN
NaN
27784      0      2      NaN      NaN
NaN
27785      0      1      SR Watson      run out  KH
Pandya
27786      0      2      NaN      NaN
NaN
27787      0      0      SN Thakur      lbw
NaN

```

```
[13565 rows x 24 columns]
```

```
delivery_df.shape
```

```
(13565, 24)
```

```

delivery_df['current_score']=delivery_df.groupby('match_id').cumsum()
['total_runs_y']
delivery_df['current_score']

```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\3255284664.py:1:
```

```
SettingWithCopyWarning:
```

```

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

```
See the caveats in the documentation:
```

```

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy

```

```
delivery_df['current_score']=delivery_df.groupby('match_id').cumsum()
['total_runs_y']
```

```
122      0
123      1
124      2
125      3
126      3
```

```
...
27783    152
27784    154
27785    155
27786    157
27787    157
```

```
Name: current_score, Length: 13565, dtype: int64
```

```
delivery_df
```

	match_id	city	winner	total_runs_x	inning
\					
122	7894	Mumbai	Chennai Super Kings	169	2
123	7894	Mumbai	Chennai Super Kings	169	2
124	7894	Mumbai	Chennai Super Kings	169	2
125	7894	Mumbai	Chennai Super Kings	169	2
126	7894	Mumbai	Chennai Super Kings	169	2
...

27783	11415	Hyderabad	Mumbai Indians	152	2
27784	11415	Hyderabad	Mumbai Indians	152	2
27785	11415	Hyderabad	Mumbai Indians	152	2
27786	11415	Hyderabad	Mumbai Indians	152	2
27787	11415	Hyderabad	Mumbai Indians	152	2

	batting_team	bowling_team	over	ball	batsman	...
\						
122	Chennai Super Kings	Mumbai Indians	1	1	SR Watson	...
123	Chennai Super Kings	Mumbai Indians	1	2	SR Watson	...
124	Chennai Super Kings	Mumbai Indians	1	3	AT Rayudu	...

125	Chennai Super Kings	Mumbai Indians	1	4	SR Watson	...
126	Chennai Super Kings	Mumbai Indians	1	5	AT Rayudu	...
...
27783	Chennai Super Kings	Mumbai Indians	20	2	RA Jadeja	...
27784	Chennai Super Kings	Mumbai Indians	20	3	SR Watson	...
27785	Chennai Super Kings	Mumbai Indians	20	4	SR Watson	...
27786	Chennai Super Kings	Mumbai Indians	20	5	SN Thakur	...
27787	Chennai Super Kings	Mumbai Indians	20	6	SN Thakur	...
	legbye_runs	noball_runs	penalty_runs	batsman_runs	extra_runs	
\						
122	0	0	0	0	0	
123	0	0	0	1	0	
124	0	0	0	1	0	
125	0	0	0	1	0	
126	0	0	0	0	0	
...	
27783	0	0	0	1	0	
27784	0	0	0	2	0	
27785	0	0	0	1	0	
27786	0	0	0	2	0	
27787	0	0	0	0	0	
	total_runs_y	player_dismissed	dismissal_kind	fielder	\	
122	0	NaN	NaN	NaN	NaN	
123	1	NaN	NaN	NaN	NaN	
124	1	NaN	NaN	NaN	NaN	
125	1	NaN	NaN	NaN	NaN	
126	0	NaN	NaN	NaN	NaN	
...	
27783	1	NaN	NaN	NaN	NaN	
27784	2	NaN	NaN	NaN	NaN	

27785	1	SR Watson	run	out	KH Pandya
27786	2	NaN		NaN	NaN
27787	0	SN Thakur	lbw		NaN

	current_score
122	0
123	1
124	2
125	3
126	3
...	...
27783	152
27784	154
27785	155
27786	157
27787	157

[13565 rows x 25 columns]

```
delivery_df['runs_left']=delivery_df['total_runs_x']-
delivery_df['current_score']
delivery_df
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\1099504436.py:1:

SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
delivery_df['runs_left']=delivery_df['total_runs_x']-
delivery_df['current_score']
```

	match_id	city	winner	total_runs_x	inning
\					
122	7894	Mumbai	Chennai Super Kings	169	2
123	7894	Mumbai	Chennai Super Kings	169	2
124	7894	Mumbai	Chennai Super Kings	169	2
125	7894	Mumbai	Chennai Super Kings	169	2
126	7894	Mumbai	Chennai Super Kings	169	2
...
27783	11415	Hyderabad	Mumbai Indians	152	2
27784	11415	Hyderabad	Mumbai Indians	152	2

27785	11415	Hyderabad	Mumbai Indians	152	2		
27786	11415	Hyderabad	Mumbai Indians	152	2		
27787	11415	Hyderabad	Mumbai Indians	152	2		
		batting_team	bowling_team	over	ball	batsman	...
\							
122	Chennai Super Kings	Mumbai Indians	1	1	SR Watson	...	
123	Chennai Super Kings	Mumbai Indians	1	2	SR Watson	...	
124	Chennai Super Kings	Mumbai Indians	1	3	AT Rayudu	...	
125	Chennai Super Kings	Mumbai Indians	1	4	SR Watson	...	
126	Chennai Super Kings	Mumbai Indians	1	5	AT Rayudu	...	
...
27783	Chennai Super Kings	Mumbai Indians	20	2	RA Jadeja	...	
27784	Chennai Super Kings	Mumbai Indians	20	3	SR Watson	...	
27785	Chennai Super Kings	Mumbai Indians	20	4	SR Watson	...	
27786	Chennai Super Kings	Mumbai Indians	20	5	SN Thakur	...	
27787	Chennai Super Kings	Mumbai Indians	20	6	SN Thakur	...	
		noball_runs	penalty_runs	batsman_runs	extra_runs	total_runs_y	
\							
122	0	0	0	0	0	0	
123	0	0	1	0	1		
124	0	0	1	0	1		
125	0	0	1	0	1		
126	0	0	0	0	0		
...
27783	0	0	1	0	1		
27784	0	0	2	0	2		
27785	0	0	1	0	1		

27786	0	0	2	0	2
27787	0	0	0	0	0

	player_dismissed	dismissal_kind	fielder	current_score
runs_left				
122	NaN	NaN	NaN	0
169				
123	NaN	NaN	NaN	1
168				
124	NaN	NaN	NaN	2
167				
125	NaN	NaN	NaN	3
166				
126	NaN	NaN	NaN	3
166				
...
...				
27783	NaN	NaN	NaN	152
0				
27784	NaN	NaN	NaN	154
-2				
27785	SR Watson	run out	KH Pandya	155
-3				
27786	NaN	NaN	NaN	157
-5				
27787	SN Thakur	lbw	NaN	157
-5				

[13565 rows x 26 columns]

```
delivery_df['balls_left']=126-
(delivery_df['over']*6+delivery_df['ball'])
delivery_df['balls_left']
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\305086741.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
delivery_df['balls_left']=126-
(delivery_df['over']*6+delivery_df['ball'])
```

122	119
123	118
124	117

```

125      116
126      115
...
27783     4
27784     3
27785     2
27786     1
27787     0

```

Name: balls_left, Length: 13565, dtype: int64

delivery_df

	match_id	city	winner	total_runs_x	inning
\					
122	7894	Mumbai	Chennai Super Kings	169	2
123	7894	Mumbai	Chennai Super Kings	169	2
124	7894	Mumbai	Chennai Super Kings	169	2
125	7894	Mumbai	Chennai Super Kings	169	2
126	7894	Mumbai	Chennai Super Kings	169	2
...
27783	11415	Hyderabad	Mumbai Indians	152	2
27784	11415	Hyderabad	Mumbai Indians	152	2
27785	11415	Hyderabad	Mumbai Indians	152	2
27786	11415	Hyderabad	Mumbai Indians	152	2
27787	11415	Hyderabad	Mumbai Indians	152	2

	batting_team	bowling_team	over	ball	batsman	...
\						
122	Chennai Super Kings	Mumbai Indians	1	1	SR Watson	...
123	Chennai Super Kings	Mumbai Indians	1	2	SR Watson	...
124	Chennai Super Kings	Mumbai Indians	1	3	AT Rayudu	...
125	Chennai Super Kings	Mumbai Indians	1	4	SR Watson	...
126	Chennai Super Kings	Mumbai Indians	1	5	AT Rayudu	...
...

27783	Chennai Super Kings	Mumbai Indians	20	2	RA Jadeja	...
27784	Chennai Super Kings	Mumbai Indians	20	3	SR Watson	...
27785	Chennai Super Kings	Mumbai Indians	20	4	SR Watson	...
27786	Chennai Super Kings	Mumbai Indians	20	5	SN Thakur	...
27787	Chennai Super Kings	Mumbai Indians	20	6	SN Thakur	...
	penalty_runs	batsman_runs	extra_runs	total_runs_y		
player_dismissed \						
122	0	0	0	0		
NaN						
123	0	1	0	1		
NaN						
124	0	1	0	1		
NaN						
125	0	1	0	1		
NaN						
126	0	0	0	0		
NaN						
...		
...						
27783	0	1	0	1		
NaN						
27784	0	2	0	2		
NaN						
27785	0	1	0	1	SR	
Watson						
27786	0	2	0	2		
NaN						
27787	0	0	0	0	SN	
Thakur						
	dismissal_kind	fielder	current_score	runs_left	balls_left	
122	NaN	NaN	0	169	119	
123	NaN	NaN	1	168	118	
124	NaN	NaN	2	167	117	
125	NaN	NaN	3	166	116	
126	NaN	NaN	3	166	115	
...
27783	NaN	NaN	152	0	4	

27784	NaN	NaN	154	-2	3
27785	run out	KH Pandya	155	-3	2
27786	NaN	NaN	157	-5	1
27787	lbw	NaN	157	-5	0

[13565 rows x 27 columns]

```
delivery_df['player_dismissed']=delivery_df['player_dismissed'].fillna('0')
delivery_df['player_dismissed']=delivery_df['player_dismissed'].apply(
lambda x:x if x == '0' else '1')
delivery_df['player_dismissed']=delivery_df['player_dismissed'].astype('int')
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\3549811476.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
delivery_df['player_dismissed']=delivery_df['player_dismissed'].fillna('0')
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\3549811476.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
delivery_df['player_dismissed']=delivery_df['player_dismissed'].apply(
lambda x:x if x == '0' else '1')
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\3549811476.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
delivery_df['player_dismissed']=delivery_df['player_dismissed'].astype('int')
```

```
wickets=delivery_df.groupby('match_id').cumsum()
['player_dismissed'].values
delivery_df['wickets']=10-wickets
delivery_df.head()
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\1442848416.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
delivery_df['wickets']=10-wickets

	match_id	city	winner	total_runs_x	inning	\
122	7894	Mumbai	Chennai Super Kings	169	2	
123	7894	Mumbai	Chennai Super Kings	169	2	
124	7894	Mumbai	Chennai Super Kings	169	2	
125	7894	Mumbai	Chennai Super Kings	169	2	
126	7894	Mumbai	Chennai Super Kings	169	2	

	batting_team	bowling_team	over	ball	
122	Chennai Super Kings	Mumbai Indians	1	1	SR Watson ...
123	Chennai Super Kings	Mumbai Indians	1	2	SR Watson ...
124	Chennai Super Kings	Mumbai Indians	1	3	AT Rayudu ...
125	Chennai Super Kings	Mumbai Indians	1	4	SR Watson ...
126	Chennai Super Kings	Mumbai Indians	1	5	AT Rayudu ...

	batman_runs	extra_runs	total_runs_y	player_dismissed
122	0	0	0	0
NaN				
123	1	0	1	0
NaN				
124	1	0	1	0
NaN				
125	1	0	1	0
NaN				
126	0	0	0	0
NaN				

	fielder	current_score	runs_left	balls_left	wickets
122	NaN	0	169	119	10
123	NaN	1	168	118	10
124	NaN	2	167	117	10
125	NaN	3	166	116	10
126	NaN	3	166	115	10

[5 rows x 28 columns]

delivery_df.tail()

	match_id	city	winner	total_runs_x	inning	\
27783	11415	Hyderabad	Mumbai Indians	152	2	
27784	11415	Hyderabad	Mumbai Indians	152	2	
27785	11415	Hyderabad	Mumbai Indians	152	2	
27786	11415	Hyderabad	Mumbai Indians	152	2	
27787	11415	Hyderabad	Mumbai Indians	152	2	

	batting_team	bowling_team	over	ball	batsman	...
\						
27783	Chennai Super Kings	Mumbai Indians	20	2	RA Jadeja	...
27784	Chennai Super Kings	Mumbai Indians	20	3	SR Watson	...
27785	Chennai Super Kings	Mumbai Indians	20	4	SR Watson	...
27786	Chennai Super Kings	Mumbai Indians	20	5	SN Thakur	...
27787	Chennai Super Kings	Mumbai Indians	20	6	SN Thakur	...

	batsman_runs	extra_runs	total_runs_y	player_dismissed
dismissal_kind				
27783	1	0	1	0
NaN				
27784	2	0	2	0
NaN				
27785	1	0	1	1
run out				
27786	2	0	2	0
NaN				
27787	0	0	0	1
lbw				

	fielder	current_score	runs_left	balls_left	wickets
27783	NaN	152	0	4	5
27784	NaN	154	-2	3	5
27785	KH Pandya	155	-3	2	4
27786	NaN	157	-5	1	4
27787	NaN	157	-5	0	3

```
[5 rows x 28 columns]
```

```
#crr=runs/overs
```

```
delivery_df['crr']=(delivery_df['current_score']*6)/(120-  
delivery_df['balls_left'])
```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\1733835164.py:3:
```

```
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation:
```

```
https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#  
returning-a-view-versus-a-copy
```

```
delivery_df['crr']=(delivery_df['current_score']*6)/(120-  
delivery_df['balls_left'])
```

```
#rrr
```

```
delivery_df['rrr']=(delivery_df['runs_left']*6)/delivery_df['balls_lef  
t']
```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\3492980221.py:2:
```

```
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation:
```

```
https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#  
returning-a-view-versus-a-copy
```

```
delivery_df['rrr']=(delivery_df['runs_left']*6)/delivery_df['balls_lef  
t']
```

```
delivery_df
```

	match_id	city	winner	total_runs_x	inning
\					
122	7894	Mumbai	Chennai Super Kings	169	2
123	7894	Mumbai	Chennai Super Kings	169	2
124	7894	Mumbai	Chennai Super Kings	169	2
125	7894	Mumbai	Chennai Super Kings	169	2
126	7894	Mumbai	Chennai Super Kings	169	2
...

27783	11415	Hyderabad	Mumbai Indians	152	2
27784	11415	Hyderabad	Mumbai Indians	152	2
27785	11415	Hyderabad	Mumbai Indians	152	2
27786	11415	Hyderabad	Mumbai Indians	152	2
27787	11415	Hyderabad	Mumbai Indians	152	2

	batting_team	bowling_team	over	ball	batsman	...
\						
122	Chennai Super Kings	Mumbai Indians	1	1	SR Watson	...
123	Chennai Super Kings	Mumbai Indians	1	2	SR Watson	...
124	Chennai Super Kings	Mumbai Indians	1	3	AT Rayudu	...
125	Chennai Super Kings	Mumbai Indians	1	4	SR Watson	...
126	Chennai Super Kings	Mumbai Indians	1	5	AT Rayudu	...
...
27783	Chennai Super Kings	Mumbai Indians	20	2	RA Jadeja	...
27784	Chennai Super Kings	Mumbai Indians	20	3	SR Watson	...
27785	Chennai Super Kings	Mumbai Indians	20	4	SR Watson	...
27786	Chennai Super Kings	Mumbai Indians	20	5	SN Thakur	...
27787	Chennai Super Kings	Mumbai Indians	20	6	SN Thakur	...

	total_runs_y	player_dismissed	dismissal_kind	fielder
current_score \				
122	0	0	NaN	NaN
0				
123	1	0	NaN	NaN
1				
124	1	0	NaN	NaN
2				
125	1	0	NaN	NaN
3				
126	0	0	NaN	NaN
3				
...
...				

27783	1	0	NaN	NaN
152				
27784	2	0	NaN	NaN
154				
27785	1	1	run out	KH Pandya
155				
27786	2	0	NaN	NaN
157				
27787	0	1	lbw	NaN
157				

	runs_left	balls_left	wickets	crr	rrr
122	169	119	10	0.000000	8.521008
123	168	118	10	3.000000	8.542373
124	167	117	10	4.000000	8.564103
125	166	116	10	4.500000	8.586207
126	166	115	10	3.600000	8.660870
...
27783	0	4	5	7.862069	0.000000
27784	-2	3	5	7.897436	-4.000000
27785	-3	2	4	7.881356	-9.000000
27786	-5	1	4	7.915966	-30.000000
27787	-5	0	3	7.850000	-inf

[13565 rows x 30 columns]

```
def result(row):
    return 1 if row['batting_team']== row['winner'] else 0
```

```
delivery_df['result']=delivery_df.apply(result,axis=1)
delivery_df
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_8420\1276910391.py:1:

SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
delivery_df['result']=delivery_df.apply(result,axis=1)
```

	match_id	city	winner	total_runs_x	inning
\					
122	7894	Mumbai	Chennai Super Kings	169	2
123	7894	Mumbai	Chennai Super Kings	169	2
124	7894	Mumbai	Chennai Super Kings	169	2
125	7894	Mumbai	Chennai Super Kings	169	2

126	0	NaN	NaN	3
166				
...
...				
27783	0	NaN	NaN	152
0				
27784	0	NaN	NaN	154
-2				
27785	1	run out	KH Pandya	155
-3				
27786	0	NaN	NaN	157
-5				
27787	1	lbw	NaN	157
-5				

	balls_left	wickets	crr	rrr	result
122	119	10	0.000000	8.521008	1
123	118	10	3.000000	8.542373	1
124	117	10	4.000000	8.564103	1
125	116	10	4.500000	8.586207	1
126	115	10	3.600000	8.660870	1
...
27783	4	5	7.862069	0.000000	0
27784	3	5	7.897436	-4.000000	0
27785	2	4	7.881356	-9.000000	0
27786	1	4	7.915966	-30.000000	0
27787	0	3	7.850000	-inf	0

[13565 rows x 31 columns]

```
final_df=delivery_df[['batting_team','bowling_team','city','runs_left',
,'balls_left','wickets','total_runs_x','crr','rrr','result']]
final_df
```

	batting_team	bowling_team	city	runs_left
balls_left \				
122	Chennai Super Kings	Mumbai Indians	Mumbai	169
119				
123	Chennai Super Kings	Mumbai Indians	Mumbai	168
118				
124	Chennai Super Kings	Mumbai Indians	Mumbai	167
117				
125	Chennai Super Kings	Mumbai Indians	Mumbai	166
116				
126	Chennai Super Kings	Mumbai Indians	Mumbai	166
115				
...
...				
27783	Chennai Super Kings	Mumbai Indians	Hyderabad	0
4				

```

27784 Chennai Super Kings Mumbai Indians Hyderabad -2
3
27785 Chennai Super Kings Mumbai Indians Hyderabad -3
2
27786 Chennai Super Kings Mumbai Indians Hyderabad -5
1
27787 Chennai Super Kings Mumbai Indians Hyderabad -5
0

```

```

      wickets  total_runs_x      crr      rrr  result
122         10          169  0.000000  8.521008        1
123         10          169  3.000000  8.542373        1
124         10          169  4.000000  8.564103        1
125         10          169  4.500000  8.586207        1
126         10          169  3.600000  8.660870        1
...         ...          ...      ...      ...      ...
27783         5          152  7.862069  0.000000        0
27784         5          152  7.897436 -4.000000        0
27785         4          152  7.881356 -9.000000        0
27786         4          152  7.915966 -30.000000        0
27787         3          152  7.850000      -inf        0

```

```
[13565 rows x 10 columns]
```

```
final_df=final_df.sample(final_df.shape[0])
```

```
final_df.sample()
```

```

      batting_team  bowling_team      city
runs_left \
15442 Royal Challengers Bangalore Mumbai Indians Bengaluru
46

```

```

      balls_left  wickets  total_runs_x      crr      rrr  result
15442         18         6          197  8.882353  15.333333        0

```

```
final_df.sample()
```

```

      batting_team  bowling_team      city  runs_left
\
10793 Royal Challengers Bangalore Kings XI Punjab Indore        61

```

```

      balls_left  wickets  total_runs_x      crr      rrr  result
10793         104        10          93  12.0  3.519231        1

```

```
final_df.isnull().sum()
```

```

batting_team    0
bowling_team    0
city            0
runs_left       0
balls_left      0
wickets         0
total_runs_x    0
crr             0
rrr            1
result          0
dtype: int64

```

```
final_df.dropna(inplace=True)
```

```
final_df = final_df[final_df['balls_left']!=0]
```

```
#final_df=pd.set_option('mode.use_inf_as_na',True)
```

```
final_df.isnull().sum()
```

```

batting_team    0
bowling_team    0
city            0
runs_left       0
balls_left      0
wickets         0
total_runs_x    0
crr             0
rrr            0
result          0
dtype: int64

```

```
final_df.head()
```

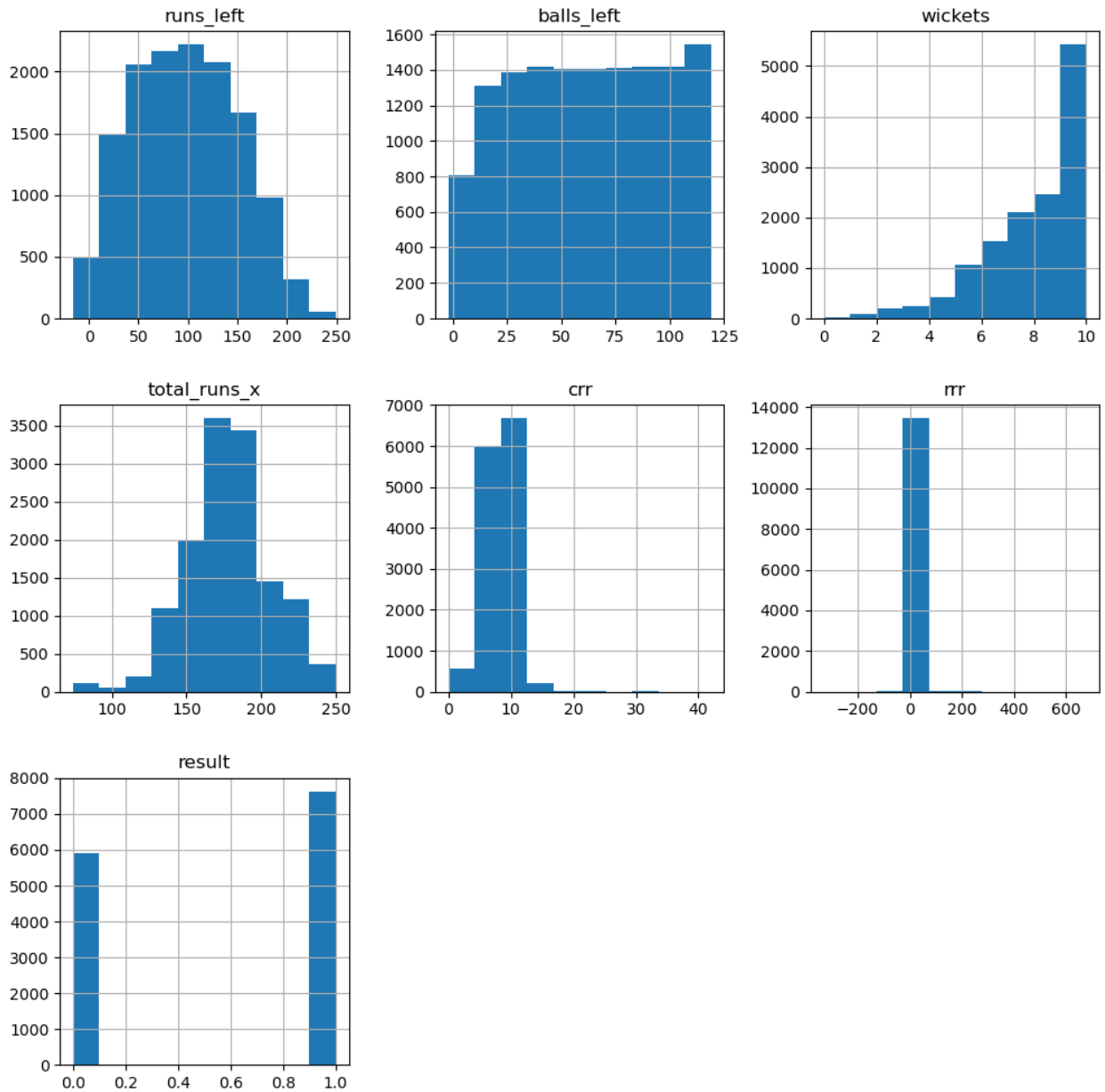
	batting_team	bowling_team
city \		
26568	Royal Challengers Bangalore	Sunrisers Hyderabad
Bengaluru		
9176	Kolkata Knight Riders	Mumbai Indians
Kolkata		
481	Kings XI Punjab	Delhi Daredevils
Mohali		
6973	Mumbai Indians	Royal Challengers Bangalore
Bengaluru		
17881	Kolkata Knight Riders	Royal Challengers Bangalore
Bengaluru		

	runs_left	balls_left	wickets	total_runs_x	crr
rrr \					
26568	49	33	7	179	8.965517
8.909091					

9176	139	51	3	217	6.782609
16.352941					
481	9	8	6	172	8.732143
6.750000					
6973	110	74	6	173	8.217391
8.918919					
17881	91	48	7	210	9.916667
11.375000					

	result
26568	1
9176	0
481	1
6973	0
17881	1

```
final_df.hist(figsize=(12,12))  
plt.show()
```



```
final_df.corr()
```

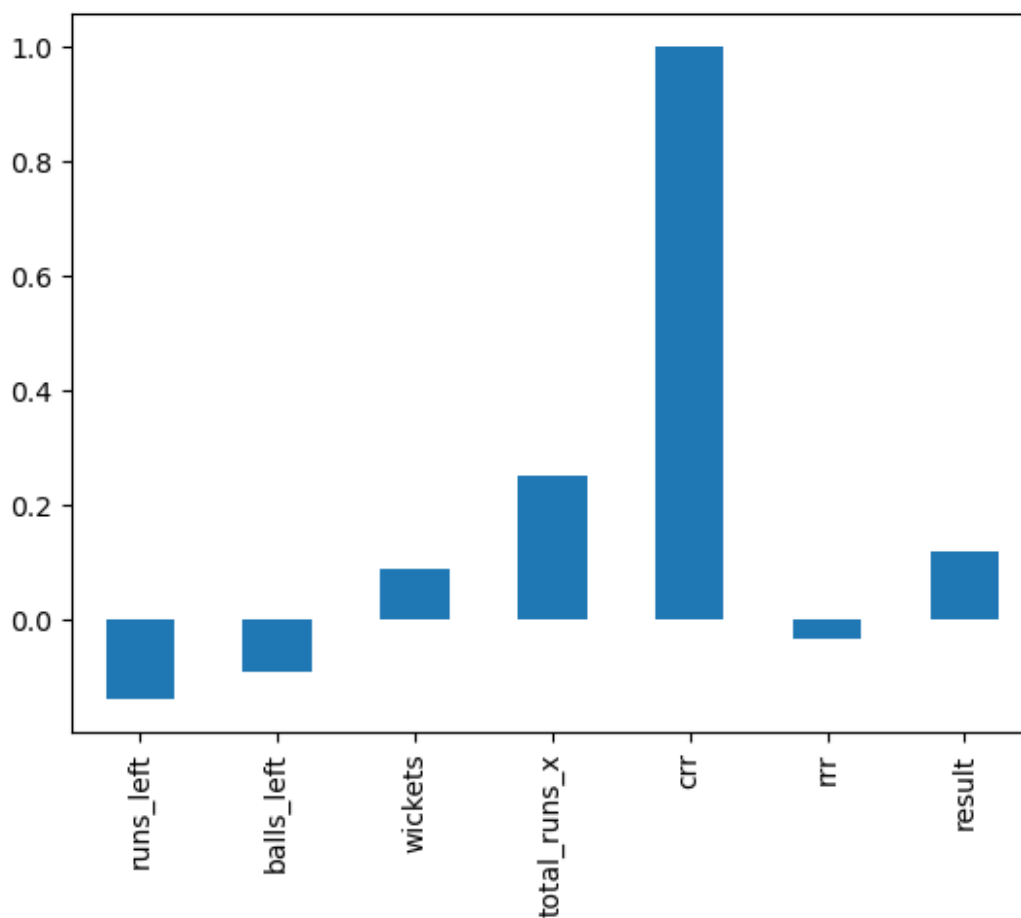
	runs_left	balls_left	wickets	total_runs_x	crr
\					
runs_left	1.000000	0.864881	0.530598	0.375743	-0.139437
balls_left	0.864881	1.000000	0.753217	-0.050943	-0.092009
wickets	0.530598	0.753217	1.000000	-0.124865	0.088336
total_runs_x	0.375743	-0.050943	-0.124865	1.000000	0.248675
crr	-0.139437	-0.092009	0.088336	0.248675	1.000000

rrr	-0.004606	-0.194362	-0.296797	0.260213	-0.033779
result	-0.218922	0.028681	0.272836	-0.350882	0.118486

	rrr	result
runs_left	-0.004606	-0.218922
balls_left	-0.194362	0.028681
wickets	-0.296797	0.272836
total_runs_x	0.260213	-0.350882
crr	-0.033779	0.118486
rrr	1.000000	-0.222523
result	-0.222523	1.000000

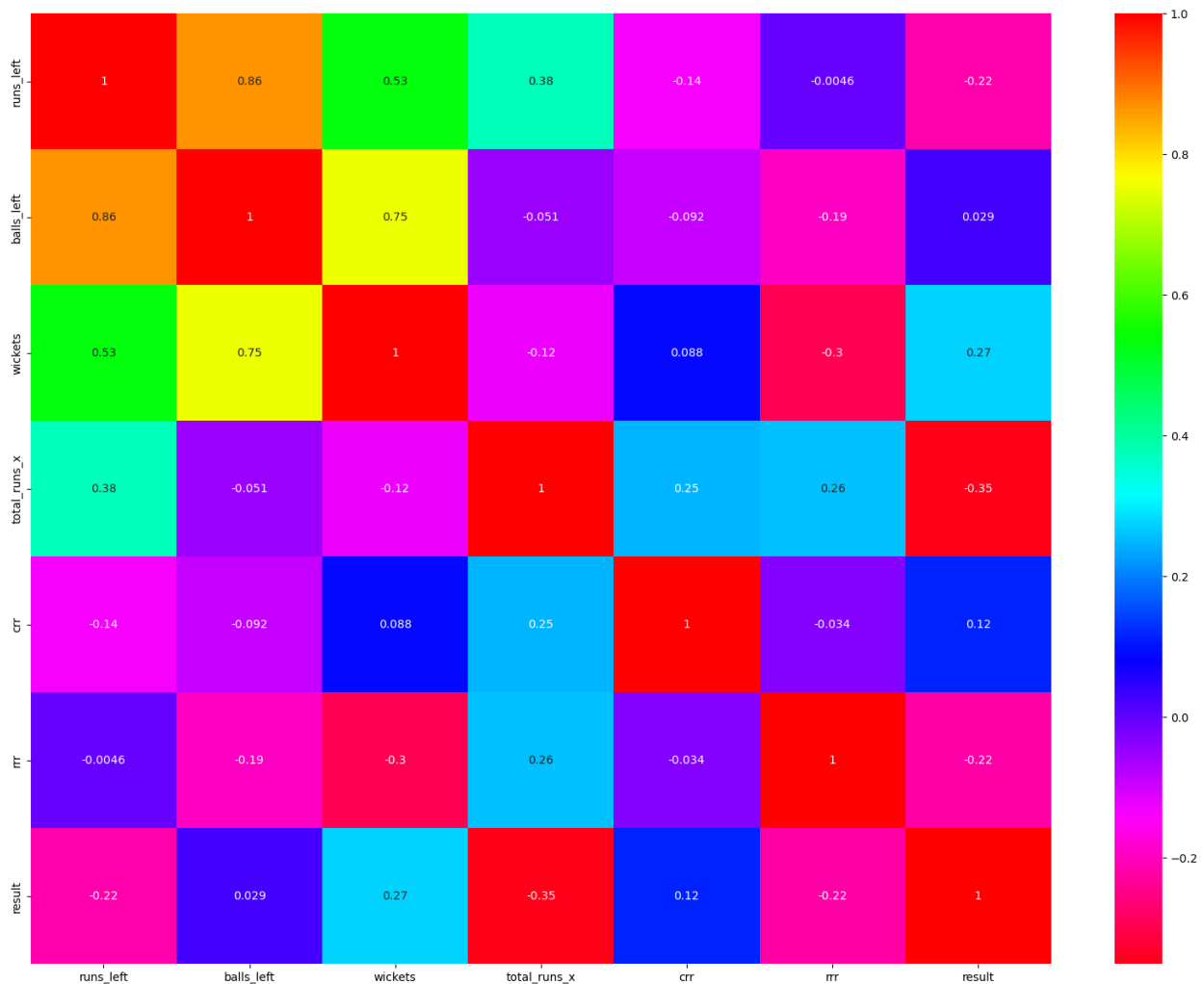
```
final_df.corr()['crr'].plot(kind='bar')
```

<AxesSubplot:>



```
plt.figure(figsize=(20,15))
corr = final_df.corr()
```

```
sns.heatmap(final_df.corr(), cmap="hsv_r", annot=True)
plt.show()
```



```
features = ['crr', 'rrr']
plt.subplots(figsize=(5,5))

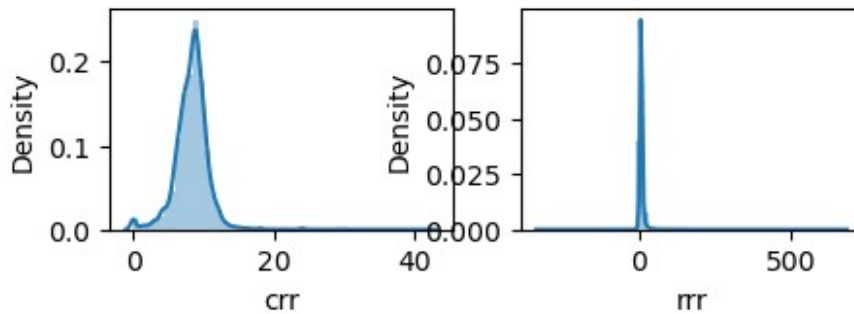
for i, col in enumerate(features):
    plt.subplot(3,2,i+1)
    sns.distplot(final_df[col])
plt.show()
```

C:\Users\Admin\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)

C:\Users\Admin\anaconda3\lib\site-packages\seaborn\

```
distributions.py:2619: FutureWarning: `distplot` is a deprecated
function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar
flexibility) or `histplot` (an axes-level function for histograms).
warnings.warn(msg, FutureWarning)
```



```
final_df=pd.get_dummies(final_df,columns=['batting_team','bowling_team',
',city'])
final_df.head()
```

	runs_left	balls_left	wickets	total_runs_x	crr
rrr \					
26568	49	33	7	179	8.965517
8.909091					
9176	139	51	3	217	6.782609
16.352941					
481	9	8	6	172	8.732143
6.750000					
6973	110	74	6	173	8.217391
8.918919					
17881	91	48	7	210	9.916667
11.375000					

	result	batting_team_Chennai	Super Kings	batting_team_Delhi
Capitals \				
26568	1		0	
0				
9176	0		0	
0				
481	1		0	
0				
6973	0		0	
0				
17881	1		0	
0				

	batting_team_Delhi	Daredevils	...	city_Chennai	city_Delhi	\
--	--------------------	------------	-----	--------------	------------	---

26568	0	...	0	0
9176	0	...	0	0
481	0	...	0	0
6973	0	...	0	0
17881	0	...	0	0

	city_Hyderabad	city_Indore	city_Jaipur	city_Kolkata
city_Mohali \				
26568	0	0	0	0
0				
9176	0	0	0	1
0				
481	0	0	0	0
1				
6973	0	0	0	0
0				
17881	0	0	0	0
0				

	city_Mumbai	city_Pune	city_Visakhapatnam
26568	0	0	0
9176	0	0	0
481	0	0	0
6973	0	0	0
17881	0	0	0

[5 rows x 36 columns]

```
x=final_df.drop('result',axis=1)
```

```
y=final_df['result']
```

```
x_train,y_train,x_test,y_test=train_test_split(x,y,test_size=0.2,random_state=1)
```

```
x_train.describe()
```

	runs_left	balls_left	wickets	total_runs_x
crr \				
count	10819.000000	10819.000000	10819.000000	10819.000000
10819.000000				
mean	97.023570	61.660412	7.611517	178.737591
8.284920				
std	53.036988	33.399660	1.988323	28.602726
2.375081				
min	-16.000000	-2.000000	0.000000	74.000000
0.000000				
25%	55.000000	33.000000	6.000000	161.000000
7.100000				
50%	96.000000	62.000000	8.000000	179.000000
8.451613				

75%	137.000000	90.000000	9.000000	196.000000
9.508954				
max	249.000000	119.000000	10.000000	250.000000
42.000000				

	rrr	batting_team_Chennai	Super Kings	\
count	10819.000000		10819.000000	
mean	10.855610		0.186709	
std	14.188550		0.389695	
min	-330.000000		0.000000	
25%	7.800000		0.000000	
50%	9.439024		0.000000	
75%	11.554374		0.000000	
max	678.000000		1.000000	

	batting_team_Delhi Capitals	batting_team_Delhi Daredevils	\
count	10819.000000	10819.000000	
mean	0.076347	0.033737	
std	0.265565	0.180560	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	0.000000	0.000000	
max	1.000000	1.000000	

	batting_team_Kings XI Punjab	...	city_Chennai	
city_Delhi	\			
count	10819.000000	...	10819.000000	10819.000000
mean	0.115815	...	0.074221	0.116092
std	0.320017	...	0.262143	0.320350
min	0.000000	...	0.000000	0.000000
25%	0.000000	...	0.000000	0.000000
50%	0.000000	...	0.000000	0.000000
75%	0.000000	...	0.000000	0.000000
max	1.000000	...	1.000000	1.000000

	city_Hyderabad	city_Indore	city_Jaipur	city_Kolkata
city_Mohali	\			
count	10819.000000	10819.000000	10819.000000	10819.000000
10819.000000				
mean	0.127738	0.030872	0.111933	0.130326
0.086330				

std	0.333814	0.172978	0.315298	0.336678
0.280863				
min	0.000000	0.000000	0.000000	0.000000
0.000000				
25%	0.000000	0.000000	0.000000	0.000000
0.000000				
50%	0.000000	0.000000	0.000000	0.000000
0.000000				
75%	0.000000	0.000000	0.000000	0.000000
0.000000				
max	1.000000	1.000000	1.000000	1.000000
1.000000				

	city_Mumbai	city_Pune	city_Visakhapatnam
count	10819.000000	10819.000000	10819.000000
mean	0.140956	0.055366	0.008966
std	0.347992	0.228703	0.094266
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000
max	1.000000	1.000000	1.000000

[8 rows x 35 columns]

y_train.describe()

	runs_left	balls_left	wickets	total_runs_x
crr \				
count	2705.000000	2705.000000	2705.000000	2705.000000
2705.000000				
mean	98.961922	63.011091	7.649908	178.778558
8.247974				
std	52.979869	33.693930	2.007518	28.613112
2.413717				
min	-12.000000	1.000000	0.000000	74.000000
0.000000				
25%	57.000000	34.000000	6.000000	161.000000
7.018868				
50%	98.000000	63.000000	8.000000	179.000000
8.457143				
75%	140.000000	93.000000	9.000000	196.000000
9.506494				
max	239.000000	119.000000	10.000000	250.000000
28.000000				

	rrr	batting_team_Chennai	Super Kings	\
count	2705.000000		2705.000000	
mean	10.811286		0.179298	
std	9.765004		0.383672	

min	-18.000000	0.000000
25%	7.852941	0.000000
50%	9.391304	0.000000
75%	11.409836	0.000000
max	234.000000	1.000000

	batting_team_Delhi Capitals	batting_team_Delhi Daredevils \
count	2705.000000	2705.000000
mean	0.078743	0.035860
std	0.269387	0.185974
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.000000	0.000000
max	1.000000	1.000000

	batting_team_Kings XI Punjab	...	city_Chennai	city_Delhi \
count	2705.000000	...	2705.000000	2705.000000
mean	0.114972	...	0.084288	0.107209
std	0.319047	...	0.277871	0.309436
min	0.000000	...	0.000000	0.000000
25%	0.000000	...	0.000000	0.000000
50%	0.000000	...	0.000000	0.000000
75%	0.000000	...	0.000000	0.000000
max	1.000000	...	1.000000	1.000000

	city_Hyderabad	city_Indore	city_Jaipur	city_Kolkata
city_Mohali \				
count	2705.000000	2705.000000	2705.000000	2705.000000
2705.000000				
mean	0.133087	0.026987	0.116451	0.126433
0.097967				
std	0.339731	0.162076	0.320824	0.332398
0.297325				
min	0.000000	0.000000	0.000000	0.000000
0.000000				
25%	0.000000	0.000000	0.000000	0.000000
0.000000				
50%	0.000000	0.000000	0.000000	0.000000
0.000000				
75%	0.000000	0.000000	0.000000	0.000000
0.000000				
max	1.000000	1.000000	1.000000	1.000000
1.000000				

	city_Mumbai	city_Pune	city_Visakhapatnam
count	2705.000000	2705.000000	2705.000000
mean	0.140111	0.043623	0.007024
std	0.347166	0.204292	0.083530
min	0.000000	0.000000	0.000000

25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000
max	1.000000	1.000000	1.000000

[8 rows x 35 columns]

```
print(x_test)
print(y_test)
```

```
10397    1
23689    1
5280     1
27200    1
10618    1
```

..

```
1885     1
9903     0
448      1
20073    1
21991    0
```

Name: result, Length: 10819, dtype: int64

```
20506    1
15876    1
2358     0
9838     0
20950    0
```

..

```
27045    1
23364    1
13181    0
22431    1
5551     0
```

Name: result, Length: 2705, dtype: int64

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test =
train_test_split(x,y,test_size=0.2,random_state=0)
```

```
from sklearn.linear_model import LogisticRegression
reg = LogisticRegression()
reg.fit(x_train,y_train)
```

C:\Users\Admin\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:814: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

```
n_iter_i = _check_optimize_result(
```

```
LogisticRegression())
```

```
y_pred=reg.predict(x_test)
```

```
from sklearn.metrics import
```

```
accuracy_score,classification_report,confusion_matrix,r2_score
```

```
print(classification_report(y_test,y_pred))
```

```
print(confusion_matrix(y_test,y_pred))
```

```
print("Training Score: ",reg.score(x_train,y_train)*100)
```

	precision	recall	f1-score	support
0	0.78	0.73	0.75	1148
1	0.81	0.85	0.83	1557
accuracy			0.80	2705
macro avg	0.80	0.79	0.79	2705
weighted avg	0.80	0.80	0.80	2705

```
[[ 837  311]
```

```
 [ 235 1322]]
```

```
Training Score: 79.08309455587393
```

```
final_df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
```

```
final_df.head()
```

	Actual	Predicted
2848	0	0
24843	0	0
8717	0	0
11777	0	1
9643	1	1

```
clfr=RandomForestClassifier(n_estimators=10,criterion='entropy',random_state=0)
```

```
clfr.fit(x_train,y_train)
```

```
RandomForestClassifier(criterion='entropy', n_estimators=10,  
random_state=0)
```

```
clfr1=RandomForestClassifier(n_estimators=10,criterion='gini',random_s  
tate=0)
```

```
clfr1.fit(x_train,y_train)
```

```
RandomForestClassifier(n_estimators=10, random_state=0)
```

```

ypre1=clfr.predict(x_test)# entropy ypre calculation
ypre2=clfr1.predict(x_test)# gini ypre calculation

print('entropy Accuracy Score:')
accuracy_score(y_test,ypre1)*100

entropy Accuracy Score:
99.9630314232902

print('gini Accuracy Score:')
accuracy_score(y_test,ypre2)*100

gini Accuracy Score:
100.0

print('entropy - confusion matrix\n-----\n')
print(confusion_matrix(y_test,ypre1))
print('gini - confusion matrix\n-----\n')
print(confusion_matrix(y_test,ypre2))

entropy - confusion matrix
-----

[[1148    0]
 [    1 1556]]
gini - confusion matrix
-----

[[1148    0]
 [    0 1557]]

print('entropy result\n-----')
print(classification_report(y_test,ypre1))
print('gini index result\n-----')
print(classification_report(y_test,ypre2))

entropy result
-----

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1148
1	1.00	1.00	1.00	1557
accuracy			1.00	2705
macro avg	1.00	1.00	1.00	2705
weighted avg	1.00	1.00	1.00	2705

```

gini index result
-----

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1148
1	1.00	1.00	1.00	1557
accuracy			1.00	2705
macro avg	1.00	1.00	1.00	2705
weighted avg	1.00	1.00	1.00	2705


```

data=pd.DataFrame({'Models':['LOGREG','RF(gini)','RF(entropy)'],
                   'Accuracy':[accuracy_score(y_test,y_pred)*100,
                                accuracy_score(y_test,ypre1)*100,accuracy_score(y_test,ypre2)*100]})
data

```

	Models	Accuracy
0	LOGREG	79.815157
1	RF(gini)	99.963031
2	RF(entropy)	100.000000