```
In [ ]:
```

```
In [1]: import numpy as np,pandas as pd,matplotlib.pyplot as plt,seaborn as sns,warnir
        warnings.filterwarnings('ignore')
```
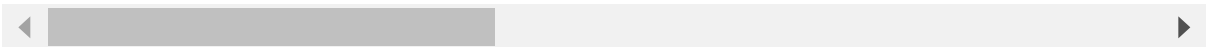
```
In [2]: d,m=pd.read_csv('deliveries.csv'),pd.read_csv('matches.csv')
```

```
In [3]: d.head()
```

Out[3]:

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_s
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	

5 rows × 21 columns

```
In [4]: m.head()
```

Out[4]:

	id	Season	city	date	team1	team2	toss_winner	toss_decision	result	dl
0	1	IPL-2017	Hyderabad	05-04-2017	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	
1	2	IPL-2017	Pune	06-04-2017	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	
2	3	IPL-2017	Rajkot	07-04-2017	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	
3	4	IPL-2017	Indore	08-04-2017	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	
4	5	IPL-2017	Bangalore	08-04-2017	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	

```
In [5]: m.shape,d.shape
```

Out[5]: ((756, 18), (179078, 21))

```
In [6]: totalruns=d.groupby(['match_id','inning']).sum()['total_runs'].reset_index()
        totalruns
```

Out[6]:

	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 × 3 columns

```
In [7]: firstinn=totalruns[totalruns['inning'] == 1]
```

```
In [8]: firstinn
```

Out[8]:

	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 × 3 columns

```
In [9]: firstinn['total_runs']=firstinn['total_runs'].map(lambda x:x+1)
        firstinn
```

Out[9]:

	match_id	inning	total_runs
0	1	1	208
2	2	1	185
4	3	1	184
6	4	1	164
8	5	1	158
...	...	...	...
1518	11347	1	144
1520	11412	1	137
1522	11413	1	172
1524	11414	1	156
1526	11415	1	153

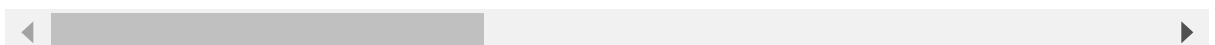756 rows × 3 columns

```
In [10]: m.columns
```

Out[10]: Index(['id', 'Season', 'city', 'date', 'team1', 'team2', 'toss_winner',
              'toss_decision', 'result', 'dl_applied', 'winner', 'win_by_runs',
              'win_by_wickets', 'player_of_match', 'venue', 'umpire1', 'umpire2',
              'umpire3'],
             dtype='object')

```
In [11]: mdf=m.merge(firstinn[['match_id','total_runs']],left_on='id',right_on='match_i
         mdf
```

Out[11]:

	id	Season	city	date	team1	team2	toss_winner	toss_decision
0	1	IPL-2017	Hyderabad	05-04-2017	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field
1	2	IPL-2017	Pune	06-04-2017	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field
2	3	IPL-2017	Rajkot	07-04-2017	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field
3	4	IPL-2017	Indore	08-04-2017	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field
4	5	IPL-2017	Bangalore	08-04-2017	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat
...	...	...	...	...	...	...	...	...
751	11347	IPL-2019	Mumbai	05-05-2019	Kolkata Knight Riders	Mumbai Indians	Mumbai Indians	field
752	11412	IPL-2019	Chennai	07-05-2019	Chennai Super Kings	Mumbai Indians	Chennai Super Kings	bat
753	11413	IPL-2019	Visakhapatnam	08-05-2019	Sunrisers Hyderabad	Delhi Capitals	Delhi Capitals	field
754	11414	IPL-2019	Visakhapatnam	10-05-2019	Delhi Capitals	Chennai Super Kings	Chennai Super Kings	field
755	11415	IPL-2019	Hyderabad	12-05-2019	Mumbai Indians	Chennai Super Kings	Mumbai Indians	bat

756 rows × 20 columns

```
In [12]: mdf['team1'].unique()

Out[12]: array(['Sunrisers Hyderabad', 'Mumbai Indians', 'Gujarat Lions',
                'Rising Pune Supergiant', 'Royal Challengers Bangalore',
                'Kolkata Knight Riders', 'Delhi Daredevils', 'Kings XI Punjab',
                'Chennai Super Kings', 'Rajasthan Royals', 'Deccan Chargers',
                'Kochi Tuskers Kerala', 'Pune Warriors', 'Rising Pune Supergiants',
                'Delhi Capitals'], dtype=object)

In [13]: teams = ['Sunrisers Hyderabad', 'Mumbai Indians','Royal Challengers Bangalore'
                  'Kolkata Knight Riders','Kings XI Punjab','Chennai Super Kings', 'Rajas

In [14]: mdf[['team1','team2']]=mdf[['team1','team2']].replace('Delhi Daredevils','Delh

In [15]: mdf=mdf[mdf['team1'].isin(teams)]
         mdf=mdf[mdf['team2'].isin(teams)]

In [16]: mdf['team1'].unique()

Out[16]: array(['Sunrisers Hyderabad', 'Royal Challengers Bangalore',
                'Kolkata Knight Riders', 'Kings XI Punjab', 'Delhi Capitals',
                'Mumbai Indians', 'Chennai Super Kings', 'Rajasthan Royals'],
               dtype=object)

In [17]: mdf=mdf[mdf['dl_applied'] == 0]

In [18]: mdf=mdf[['match_id','city','winner','total_runs']]
```

```
In [19]: mdf
```

Out[19]:

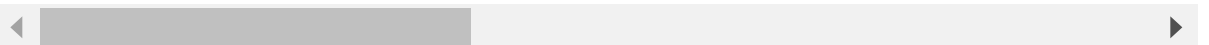|	match_id	city	winner	total_runs
0	1	Hyderabad	Sunrisers Hyderabad	208
4	5	Bangalore	Royal Challengers Bangalore	158
6	7	Mumbai	Mumbai Indians	179
7	8	Indore	Kings XI Punjab	149
9	10	Mumbai	Mumbai Indians	159
...	...	...	...	...
751	11347	Mumbai	Mumbai Indians	144
752	11412	Chennai	Mumbai Indians	137
753	11413	Visakhapatnam	Delhi Capitals	172
754	11414	Visakhapatnam	Chennai Super Kings	156
755	11415	Hyderabad	Mumbai Indians	153

626 rows × 4 columns

```
In [20]: delivery_df=mdf.merge(d,on='match_id')
```
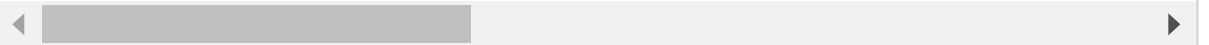
```
In [21]: delivery_df
```

Out[21]:

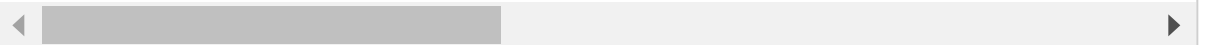|	match_id	city	winner	total_runs_x	inning	batting_team	bowling_team	over
0	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
1	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
2	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
3	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
4	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
...	...	...	...	...	...	...	...	..
149573	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149574	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149575	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149576	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149577	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20

149578 rows × 24 columns

```
In [22]: delivery_df[['batting_team','bowling_team','winner']]=delivery_df[['batting_te
```

```
In [23]: #total_run_x is the score of 1st inning whereas total_run_y is the score of se
         ##Current score of particular match
         delivery_df['current_score'] = delivery_df.groupby('match_id').cumsum()['total
```
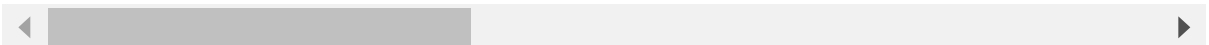
```
In [24]: delivery_df
```

Out[24]:

	match_id	city	winner	total_runs_x	inning	batting_team	bowling_team	ove
0	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
1	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
2	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
3	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
4	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
...	...	...	...	...	...	...	...	..
149573	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149574	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149575	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149576	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149577	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20

149578 rows × 25 columns

```
In [25]: #runs left in the match
         delivery_df['runs_left'] = delivery_df['total_runs_x'] - delivery_df['current_
```
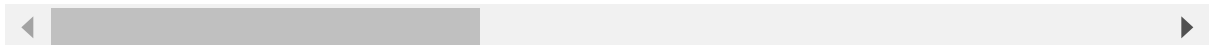
```
In [26]: delivery_df
```

Out[26]:

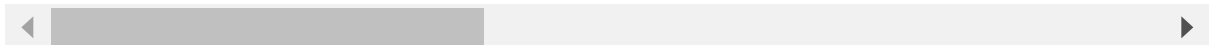|	match_id	city	winner	total_runs_x	inning	batting_team	bowling_team	over
0	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
1	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
2	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
3	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
4	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
...	...	...	...	...	...	...	...	..
149573	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149574	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149575	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149576	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149577	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20

149578 rows × 26 columns

```
In [27]: delivery_df['balls_left'] = 126 -(delivery_df['over']*6+delivery_df['ball'])
```

In [28]: delivery_df

Out[28]:

	match_id	city	winner	total_runs_x	inning	batting_team	bowling_team	over
0	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
1	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
2	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
3	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
4	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
...	...	...	...	...	...	...	...	..
149573	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149574	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149575	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149576	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149577	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20

149578 rows × 27 columns

```python
In [29]: delivery_df['player_dismissed'].unique()
```

```
Out[29]: array([nan, 'DA Warner', 'S Dhawan', 'MC Henriques', 'Yuvraj Singh',
        'Mandeep Singh', 'CH Gayle', 'KM Jadhav', 'TM Head', 'Sachin Baby',
        'STR Binny', 'SR Watson', 'S Aravind', 'TS Mills', 'YS Chahal',
        'Vishnu Vinod', 'P Negi', 'AP Tare', 'KK Nair', 'SW Billings',
        'SV Samson', 'CH Morris', 'CR Brathwaite', 'PJ Cummins', 'RR Pant',
        'S Nadeem', 'G Gambhir', 'RV Uthappa', 'CA Lynn', 'YK Pathan',
        'SA Yadav', 'CR Woakes', 'SP Narine', 'PA Patel', 'JC Buttler',
        'RG Sharma', 'KH Pandya', 'KA Pollard', 'N Rana', 'M Vohra',
        'AR Patel', 'DJ Hooda', 'BCJ Cutting', 'V Shankar', 'NV Ojha',
        'Rashid Khan', 'MP Stoinis', 'HM Amla', 'GJ Maxwell', 'DA Miller',
        'WP Saha', 'MM Sharma', 'VR Aaron', 'V Kohli', 'AB de Villiers',
        'MJ McClenaghan', 'MK Pandey', 'C de Grandhomme', 'SS Iyer',
        'EJG Morgan', 'KC Cariappa', 'AD Mathews', 'Mohammad Nabi',
        'I Sharma', 'KS Williamson', 'SE Marsh', 'Harbhajan Singh',
        'HH Pandya', 'CJ Anderson', 'K Rabada', 'NM Coulter-Nile',
        'Kuldeep Yadav', 'S Badree', 'MJ Guptill', 'Anureet Singh',
        'Mohammed Shami', 'SP Jackson', 'KV Sharma', 'A Choudhary',
        'LMP Simmons', 'MN Samuels', 'Z Khan', 'Swapnil Singh',
        'SS Tiwary', 'AT Rayudu', 'A Mishra', 'IR Jaggi', 'PP Chawla',
        'AS Rajpoot', 'SC Ganguly', 'RT Ponting', 'DJ Hussey', 'R Dravid',
        'JH Kallis', 'W Jaffer', 'MV Boucher', 'B Akhil', 'CL White',
        'AA Noffke', 'SB Joshi', 'ML Hayden', 'MS Dhoni', 'SK Raina',
        'JDP Oram', 'K Goel', 'JR Hopes', 'KC Sangakkara', 'T Kohli',
        'DS Lehmann', 'M Kaif', 'M Rawat', 'RA Jadeja', 'SK Warne',
        'V Sehwag', 'L Ronchi', 'DJ Thornely', 'ST Jayasuriya', 'PR Shah',
        'AM Nayar', 'SM Pollock', 'S Chanderpaul', 'LRPL Taylor',
        'Y Venugopal Rao', 'VVS Laxman', 'AC Gilchrist', 'SB Styris',
        'AS Yadav', 'A Symonds', 'WPUJC Vaas', 'SB Bangar', 'PP Ojha',
        'BB McCullum', 'Mohammad Hafeez', 'DPMD Jayawardene', 'IK Pathan',
        'B Lee', 'S Sohal', 'Kamran Akmal', 'Shahid Afridi', 'MEK Hussey',
        'DJ Bravo', 'MA Khote', 'GC Smith', 'D Salunkhe', 'RR Sarwan',
        'S Sreesanth', 'VRV Singh', 'P Kumar', 'AB Agarkar', 'M Kartik',
        'LR Shukla', 'AM Rahane', 'Shoaib Malik', 'KD Karthik', 'R Bhatia',
        'MK Tiwary', 'MF Maharoof', 'SM Katich', 'TM Srivastava',
        'B Chipli', 'DW Steyn', 'DB Das', 'HH Gibbs', 'DNT Zoysa',
        'D Kalyankrishna', 'SA Asnodkar', 'Salman Butt', 'BJ Hodge',
        'Umar Gul', 'AB Dinda', 'SP Fleming', 'S Vidyut', 'JA Morkel',
        'LPC Silva', 'DB Ravi Teja', 'RP Singh', 'Misbah-ul-Haq',
        'YV Takawale', 'PJ Sangwan', 'Mohammad Asif', 'GD McGrath',
        'S Badrinath', 'Joginder Sharma', 'M Muralitharan', 'M Ntini',
        'DT Patil', 'A Kumble', 'S Anirudha', 'Sohail Tanvir',
        'SK Trivedi', 'CK Kapugedera', 'A Chopra', 'T Taibu',
        'J Arunkumar', 'SP Goswami', 'R Vinay Kumar', 'AD Mascarenhas',
        'TM Dilshan', 'VY Mahesh', 'SR Tendulkar', 'Younis Khan',
        'Niraj Patel', 'Shoaib Akhtar', 'H Das', 'DR Smith', 'SD Chitnis',
        'A Nehra', 'VS Yeligati', 'MS Gony', 'L Balaji', 'LA Pomersbach',
        'A Nel', 'A Mukund', 'WA Mota', 'JP Duminy', 'A Flintoff',
        'JD Ryder', 'KP Pietersen', 'T Henderson', 'MM Patel',
        'Kamran Khan', 'R Bishnoi', 'PC Valthaty', 'RJ Quiney',
        'Yashpal Singh', 'Pankaj Singh', 'RS Bopara', 'FH Edwards',
        'AS Raut', 'AN Ghosh', 'BAW Mendis', 'DL Vettori',
        'RE van der Merwe', 'RR Powar', 'AA Bilakhia', 'TL Suman',
        'Shoaib Ahmed', 'GR Napier', 'MN van Wyk', 'LA Carseldine',
        'M Vijay', 'SB Jakati', 'M Manhas', 'RJ Harris', 'D du Preez',
        'RR Raje', 'DS Kulkarni', 'M Morkel', 'S Tyagi', 'C Nanda',
        'SL Malinga', 'J Botha', 'A Singh', 'GJ Bailey', 'AB McDonald',
        'Y Nagar', 'SS Shaikh', 'R Ashwin', 'T Thushara',
```

```
             'Mohammad Ashraful', 'CA Pujara', 'Anirudh Singh', 'R Sathish',
             'AA Jhunjhunwala', 'P Dogra', 'A Uniyal', 'MS Bisla', 'JM Kemp',
             'OA Shah', 'RS Gavaskar', 'SE Bond', 'DP Nannes', 'MJ Lumb',
             'DR Martyn', 'S Narwal', 'AB Barath', 'FY Fazal', 'MD Mishra',
             'S Ladda', 'J Theron', 'SJ Srivastava', 'R Sharma', 'AC Voges',
             'Jaskaran Singh', 'KAJ Roach', 'PD Collingwood', 'CK Langeveldt',
             'VS Malik', 'SW Tait', 'A Mithun', 'AP Dole', 'Harmeet Singh',
             'R McLaren', 'S Sriram', 'KP Appanna', 'C Madan', 'AG Paunikar',
             'MR Marsh', 'AJ Finch', 'B Sumanth', 'DT Christian', 'UBT Chand',
             'UT Yadav', 'DJ Jacobs', 'AL Menaria', 'MA Agarwal', 'AUK Pathan',
             'Sunny Singh', 'JJ van der Wath', 'R Ninan', 'S Randiv',
             'BJ Haddin', 'MS Wade', 'RN ten Doeschate', 'TR Birt',
             'Bipul Sharma', 'AC Blizzard', 'CA Ingram', 'DH Yagnik',
             'KB Arun Karthik', 'BA Bhatt', 'Shakib Al Hasan', 'JEC Franklin',
             'F du Plessis', 'DE Bollinger', 'RE Levi', 'HV Patel', 'DJ Harris',
             'Ankit Sharma', 'KK Cooper', 'AA Chavan', 'GB Hogg', 'RR Bhatkal',
             'CJ McKay', 'N Saini', 'Azhar Mahmood', 'RJ Peterson',
             'A Ashish Reddy', 'V Pratap Singh', 'Gurkeerat Singh', 'PA Reddy',
             'AD Russell', 'Sunny Gupta', 'MC Juneja', 'MDKJ Perera',
             'B Laughlin', 'NLTC Perera', 'BMAJ Mendis', 'GH Vihari',
             'R Rampaul', 'P Awana', 'SMSM Senanayake', 'BJ Rohrer', 'KL Rahul',
             'J Syed Mohammad', 'R Dhawan', 'MG Johnson', 'Q de Kock',
             'BB Samantray', 'JP Faulkner', 'X Thalaivan Sargunam', 'DJG Sammy',
             'CM Gautam', 'NJ Maddinson', 'JDS Neesham', 'B Kumar', 'SPD Smith',
             'TG Southee', 'PV Tambe', 'S Rana', 'MA Starc', 'BR Dunk',
             'RR Rossouw', 'KW Richardson', 'Shivam Sharma', 'VH Zol',
             'Imran Tahir', 'R Shukla', 'WD Parnell', 'BE Hendricks',
             'M de Lange', 'S Gopal', 'R Tewatia', 'JO Holder', 'JD Unadkat',
             'SA Abbott', 'AN Ahmed', 'Iqbal Abdulla', 'SN Khan', 'D Wiese',
             'DJ Muthuswami', 'Karanveer Singh', 'Sandeep Sharma', 'J Suchith',
             'C Munro', 'P Sahu', 'NS Naik', 'Parvez Rasool', 'JJ Bumrah',
             'J Yadav', 'CJ Jordan', 'E Lewis', 'Ishan Kishan', 'DL Chahar',
             'M Wood', 'R Singh', 'D Short', 'BA Stokes', 'RA Tripathi',
             'K Gowtham', 'S Kaul', 'AJ Tye', 'M Ur Rahman', 'S Gill', 'S Mavi',
             'Washington Sundar', 'T Curran', 'TA Boult', 'H Klaasen', 'JJ Roy',
             'R Bhui', 'J Archer', 'P Shaw', 'L Plunkett', 'Basil Thampi',
             'M Markande', 'Mustafizur Rahman', 'BB Sran', 'A Hales',
             'M Lomror', 'M Ashwin', 'Mohammed Siraj', 'D Shorey', 'M Ali',
             'AD Nath', 'P Chopra', 'I Sodhi', 'J Searles', 'S Hetmyer',
             'S Dube', 'J Bairstow', 'C Ingram', 'K Paul', 'N Pooran', 'N Naik',
             'H Vihari', 'P R Barman', 'S Curran', 'H Viljoen', 'S Lamichhane',
             'RD Chahar', 'P Krishna', 'SD Lad', 'R Parag', 'J Denly',
             'L Livingstone', 'K Ahmed', 'A Turner', 'SN Thakur',
             'S Rutherford', 'M Santner', 'S Singh'], dtype=object)
```

In [30]: `delivery_df['player_dismissed'] = delivery_df['player_dismissed'].fillna(0)`

```
In [31]: delivery_df['player_dismissed'].unique()
```

```
Out[31]: array([0, 'DA Warner', 'S Dhawan', 'MC Henriques', 'Yuvraj Singh',
       'Mandeep Singh', 'CH Gayle', 'KM Jadhav', 'TM Head', 'Sachin Baby',
       'STR Binny', 'SR Watson', 'S Aravind', 'TS Mills', 'YS Chahal',
       'Vishnu Vinod', 'P Negi', 'AP Tare', 'KK Nair', 'SW Billings',
       'SV Samson', 'CH Morris', 'CR Brathwaite', 'PJ Cummins', 'RR Pant',
       'S Nadeem', 'G Gambhir', 'RV Uthappa', 'CA Lynn', 'YK Pathan',
       'SA Yadav', 'CR Woakes', 'SP Narine', 'PA Patel', 'JC Buttler',
       'RG Sharma', 'KH Pandya', 'KA Pollard', 'N Rana', 'M Vohra',
       'AR Patel', 'DJ Hooda', 'BCJ Cutting', 'V Shankar', 'NV Ojha',
       'Rashid Khan', 'MP Stoinis', 'HM Amla', 'GJ Maxwell', 'DA Miller',
       'WP Saha', 'MM Sharma', 'VR Aaron', 'V Kohli', 'AB de Villiers',
       'MJ McClenaghan', 'MK Pandey', 'C de Grandhomme', 'SS Iyer',
       'EJG Morgan', 'KC Cariappa', 'AD Mathews', 'Mohammad Nabi',
       'I Sharma', 'KS Williamson', 'SE Marsh', 'Harbhajan Singh',
       'HH Pandya', 'CJ Anderson', 'K Rabada', 'NM Coulter-Nile',
       'Kuldeep Yadav', 'S Badree', 'MJ Guptill', 'Anureet Singh',
       'Mohammed Shami', 'SP Jackson', 'KV Sharma', 'A Choudhary',
       'LMP Simmons', 'MN Samuels', 'Z Khan', 'Swapnil Singh',
       'SS Tiwary', 'AT Rayudu', 'A Mishra', 'IR Jaggi', 'PP Chawla',
       'AS Rajpoot', 'SC Ganguly', 'RT Ponting', 'DJ Hussey', 'R Dravid',
       'JH Kallis', 'W Jaffer', 'MV Boucher', 'B Akhil', 'CL White',
       'AA Noffke', 'SB Joshi', 'ML Hayden', 'MS Dhoni', 'SK Raina',
       'JDP Oram', 'K Goel', 'JR Hopes', 'KC Sangakkara', 'T Kohli',
       'DS Lehmann', 'M Kaif', 'M Rawat', 'RA Jadeja', 'SK Warne',
       'V Sehwag', 'L Ronchi', 'DJ Thornely', 'ST Jayasuriya', 'PR Shah',
       'AM Nayar', 'SM Pollock', 'S Chanderpaul', 'LRPL Taylor',
       'Y Venugopal Rao', 'VVS Laxman', 'AC Gilchrist', 'SB Styris',
       'AS Yadav', 'A Symonds', 'WPUJC Vaas', 'SB Bangar', 'PP Ojha',
       'BB McCullum', 'Mohammad Hafeez', 'DPMD Jayawardene', 'IK Pathan',
       'B Lee', 'S Sohal', 'Kamran Akmal', 'Shahid Afridi', 'MEK Hussey',
       'DJ Bravo', 'MA Khote', 'GC Smith', 'D Salunkhe', 'RR Sarwan',
       'S Sreesanth', 'VRV Singh', 'P Kumar', 'AB Agarkar', 'M Kartik',
       'LR Shukla', 'AM Rahane', 'Shoaib Malik', 'KD Karthik', 'R Bhatia',
       'MK Tiwary', 'MF Maharoof', 'SM Katich', 'TM Srivastava',
       'B Chipli', 'DW Steyn', 'DB Das', 'HH Gibbs', 'DNT Zoysa',
       'D Kalyankrishna', 'SA Asnodkar', 'Salman Butt', 'BJ Hodge',
       'Umar Gul', 'AB Dinda', 'SP Fleming', 'S Vidyut', 'JA Morkel',
       'LPC Silva', 'DB Ravi Teja', 'RP Singh', 'Misbah-ul-Haq',
       'YV Takawale', 'PJ Sangwan', 'Mohammad Asif', 'GD McGrath',
       'S Badrinath', 'Joginder Sharma', 'M Muralitharan', 'M Ntini',
       'DT Patil', 'A Kumble', 'S Anirudha', 'Sohail Tanvir',
       'SK Trivedi', 'CK Kapugedera', 'A Chopra', 'T Taibu',
       'J Arunkumar', 'SP Goswami', 'R Vinay Kumar', 'AD Mascarenhas',
       'TM Dilshan', 'VY Mahesh', 'SR Tendulkar', 'Younis Khan',
       'Niraj Patel', 'Shoaib Akhtar', 'H Das', 'DR Smith', 'SD Chitnis',
       'A Nehra', 'VS Yeligati', 'MS Gony', 'L Balaji', 'LA Pomersbach',
       'A Nel', 'A Mukund', 'WA Mota', 'JP Duminy', 'A Flintoff',
       'JD Ryder', 'KP Pietersen', 'T Henderson', 'MM Patel',
       'Kamran Khan', 'R Bishnoi', 'PC Valthaty', 'RJ Quiney',
       'Yashpal Singh', 'Pankaj Singh', 'RS Bopara', 'FH Edwards',
       'AS Raut', 'AN Ghosh', 'BAW Mendis', 'DL Vettori',
       'RE van der Merwe', 'RR Powar', 'AA Bilakhia', 'TL Suman',
       'Shoaib Ahmed', 'GR Napier', 'MN van Wyk', 'LA Carseldine',
       'M Vijay', 'SB Jakati', 'M Manhas', 'RJ Harris', 'D du Preez',
       'RR Raje', 'DS Kulkarni', 'M Morkel', 'S Tyagi', 'C Nanda',
       'SL Malinga', 'J Botha', 'A Singh', 'GJ Bailey', 'AB McDonald',
       'Y Nagar', 'SS Shaikh', 'R Ashwin', 'T Thushara',
```

```
       'Mohammad Ashraful', 'CA Pujara', 'Anirudh Singh', 'R Sathish',
       'AA Jhunjhunwala', 'P Dogra', 'A Uniyal', 'MS Bisla', 'JM Kemp',
       'OA Shah', 'RS Gavaskar', 'SE Bond', 'DP Nannes', 'MJ Lumb',
       'DR Martyn', 'S Narwal', 'AB Barath', 'FY Fazal', 'MD Mishra',
       'S Ladda', 'J Theron', 'SJ Srivastava', 'R Sharma', 'AC Voges',
       'Jaskaran Singh', 'KAJ Roach', 'PD Collingwood', 'CK Langeveldt',
       'VS Malik', 'SW Tait', 'A Mithun', 'AP Dole', 'Harmeet Singh',
       'R McLaren', 'S Sriram', 'KP Appanna', 'C Madan', 'AG Paunikar',
       'MR Marsh', 'AJ Finch', 'B Sumanth', 'DT Christian', 'UBT Chand',
       'UT Yadav', 'DJ Jacobs', 'AL Menaria', 'MA Agarwal', 'AUK Pathan',
       'Sunny Singh', 'JJ van der Wath', 'R Ninan', 'S Randiv',
       'BJ Haddin', 'MS Wade', 'RN ten Doeschate', 'TR Birt',
       'Bipul Sharma', 'AC Blizzard', 'CA Ingram', 'DH Yagnik',
       'KB Arun Karthik', 'BA Bhatt', 'Shakib Al Hasan', 'JEC Franklin',
       'F du Plessis', 'DE Bollinger', 'RE Levi', 'HV Patel', 'DJ Harris',
       'Ankit Sharma', 'KK Cooper', 'AA Chavan', 'GB Hogg', 'RR Bhatkal',
       'CJ McKay', 'N Saini', 'Azhar Mahmood', 'RJ Peterson',
       'A Ashish Reddy', 'V Pratap Singh', 'Gurkeerat Singh', 'PA Reddy',
       'AD Russell', 'Sunny Gupta', 'MC Juneja', 'MDKJ Perera',
       'B Laughlin', 'NLTC Perera', 'BMAJ Mendis', 'GH Vihari',
       'R Rampaul', 'P Awana', 'SMSM Senanayake', 'BJ Rohrer', 'KL Rahul',
       'J Syed Mohammad', 'R Dhawan', 'MG Johnson', 'Q de Kock',
       'BB Samantray', 'JP Faulkner', 'X Thalaivan Sargunam', 'DJG Sammy',
       'CM Gautam', 'NJ Maddinson', 'JDS Neesham', 'B Kumar', 'SPD Smith',
       'TG Southee', 'PV Tambe', 'S Rana', 'MA Starc', 'BR Dunk',
       'RR Rossouw', 'KW Richardson', 'Shivam Sharma', 'VH Zol',
       'Imran Tahir', 'R Shukla', 'WD Parnell', 'BE Hendricks',
       'M de Lange', 'S Gopal', 'R Tewatia', 'JO Holder', 'JD Unadkat',
       'SA Abbott', 'AN Ahmed', 'Iqbal Abdulla', 'SN Khan', 'D Wiese',
       'DJ Muthuswami', 'Karanveer Singh', 'Sandeep Sharma', 'J Suchith',
       'C Munro', 'P Sahu', 'NS Naik', 'Parvez Rasool', 'JJ Bumrah',
       'J Yadav', 'CJ Jordan', 'E Lewis', 'Ishan Kishan', 'DL Chahar',
       'M Wood', 'R Singh', 'D Short', 'BA Stokes', 'RA Tripathi',
       'K Gowtham', 'S Kaul', 'AJ Tye', 'M Ur Rahman', 'S Gill', 'S Mavi',
       'Washington Sundar', 'T Curran', 'TA Boult', 'H Klaasen', 'JJ Roy',
       'R Bhui', 'J Archer', 'P Shaw', 'L Plunkett', 'Basil Thampi',
       'M Markande', 'Mustafizur Rahman', 'BB Sran', 'A Hales',
       'M Lomror', 'M Ashwin', 'Mohammed Siraj', 'D Shorey', 'M Ali',
       'AD Nath', 'P Chopra', 'I Sodhi', 'J Searles', 'S Hetmyer',
       'S Dube', 'J Bairstow', 'C Ingram', 'K Paul', 'N Pooran', 'N Naik',
       'H Vihari', 'P R Barman', 'S Curran', 'H Viljoen', 'S Lamichhane',
       'RD Chahar', 'P Krishna', 'SD Lad', 'R Parag', 'J Denly',
       'L Livingstone', 'K Ahmed', 'A Turner', 'SN Thakur',
       'S Rutherford', 'M Santner', 'S Singh'], dtype=object)
```

In [32]: `delivery_df['player_dismissed'] = np.where(delivery_df['player_dismissed'] ==`

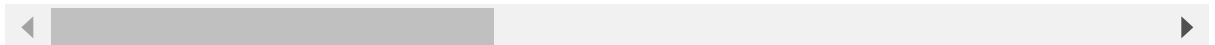In [33]: `delivery_df['player_dismissed'].unique()`

Out[33]: `array([0, 1])`

```
In [34]: delivery_df['wickets'] = 10 - (delivery_df.groupby('match_id').cumsum()['playe
```

```
In [35]: delivery_df
```

Out[35]:

	match_id	city	winner	total_runs_x	inning	batting_team	bowling_team	over
0	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
1	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
2	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
3	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
4	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
...	...	...	...	...	...	...	...	..
149573	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149574	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149575	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149576	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20
149577	11415	Hyderabad	Mumbai Indians	153	2	Chennai Super Kings	Mumbai Indians	20

149578 rows × 28 columns

```
In [36]: delivery_df['curr_run_rate'] = (delivery_df['current_score']*6)/(120 - deliver
```
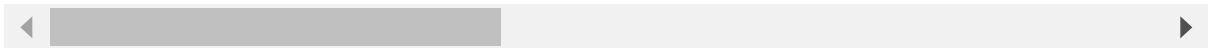
```
In [37]: delivery_df['req_run_rate'] = (delivery_df['runs_left']*6)/(delivery_df['balls
```

```
In [38]:   delivery_df.head()
```

Out[38]:

	match_id	city	winner	total_runs_x	inning	batting_team	bowling_team	over	ball
0	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1
1	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2
2	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3
3	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4
4	1	Hyderabad	Sunrisers Hyderabad	208	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5

5 rows × 30 columns

```
In [39]:   # def result(row):
           #     return 1 if row['batting_team'] == row['winner'] else 0

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

           #################OR

           delivery_df['result'] = np.where(delivery_df['batting_team'] == delivery_df['w
```
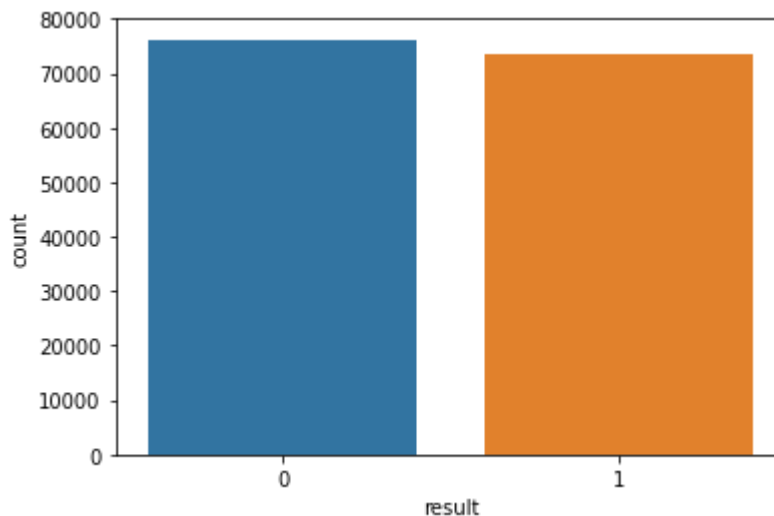
```
In [40]: sns.countplot(delivery_df['result'])
         delivery_df['result'].value_counts()
```

Out[40]: 0   76215
         1   73363
         Name: result, dtype: int64


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

Out[41]: Index(['match_id', 'city', 'winner', 'total_runs_x', 'inning', 'batting_tea
         m',
                'bowling_team', 'over', 'ball', 'batsman', 'non_striker', 'bowler',
                'is_super_over', 'wide_runs', 'bye_runs', 'legbye_runs', 'noball_run
         s',
                'penalty_runs', 'batsman_runs', 'extra_runs', 'total_runs_y',
                'player_dismissed', 'dismissal_kind', 'fielder', 'current_score',
                'runs_left', 'balls_left', 'wickets', 'curr_run_rate', 'req_run_rat
         e',
                'result'],
               dtype='object')
```

```
In [42]: final_df = delivery_df[['batting_team','bowling_team','city','total_runs_x','
```

```
In [43]: final_df = final_df[final_df['balls_left'] != 0]
```

```
In [44]: final_df.dropna(inplace=True)
```

```
In [45]: final_df[(final_df['req_run_rate']!=0)]
```

Out[45]:

	batting_team	bowling_team	city	total_runs_x	curr_run_rate	req_run_rate	wicke
0	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	0.000000	10.487395	
1	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	0.000000	10.576271	
2	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	8.000000	10.461538	
3	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	6.000000	10.551724	
4	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	7.200000	10.539130	
...	...	...	...	...	...	...	
149572	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.808696	-180.000000	
149573	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.724138	-226.500000	
149574	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.692308	-306.000000	
149575	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.610169	-462.000000	
149576	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.579832	-936.000000	

146260 rows × 9 columns

```
In [46]: final_df.result.value_counts()
```

```
Out[46]: 0    74838
         1    72225
         Name: result, dtype: int64
```

```
In [47]: final_df.city.unique()
```

```
Out[47]: array(['Hyderabad', 'Bangalore', 'Mumbai', 'Indore', 'Kolkata', 'Delhi',
                'Chandigarh', 'Jaipur', 'Chennai', 'Cape Town', 'Port Elizabeth',
                'Durban', 'Centurion', 'East London', 'Johannesburg', 'Kimberley',
                'Bloemfontein', 'Ahmedabad', 'Cuttack', 'Nagpur', 'Dharamsala',
                'Visakhapatnam', 'Pune', 'Raipur', 'Ranchi', 'Abu Dhabi',
                'Sharjah', 'Mohali', 'Bengaluru'], dtype=object)
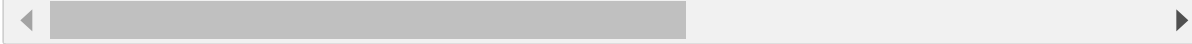```

```
In [48]: # notdone = final_df[final_df["result"] == 0]
         # done = final_df[final_df["result"] == 1]

         # # Resampling the original dataset with 1,95,482 datapoints for both classes
         # notdone_resample = notdone.sample(n = 72225, random_state = 193)
         # done_resample = done.sample(n = 72225, random_state = 193)
         # df_eda=final_df.copy()
         # # Creating new dataset consisting of equal class occurence
         # df = pd.concat([notdone_resample, done_resample], axis = 0)
```

```
In [49]: x = final_df.drop('result',1)
         y = final_df['result']
```

```
In [50]: from sklearn.compose import ColumnTransformer
         from sklearn.preprocessing import OneHotEncoder
         from sklearn.pipeline import Pipeline
```

```
In [51]: c = final_df.select_dtypes(include='O').columns
         cf=ColumnTransformer([('trf',OneHotEncoder(sparse=False,drop='first'),['battir
```

```
In [52]: # c = final_df.select_dtypes(include='O').columns
         # from sklearn.preprocessing import LabelEncoder
         # le = LabelEncoder()
         # for i in c:
         #     x[i] = le.fit_transform(x[i])
```

In [53]: `x`

Out[53]:

	batting_team	bowling_team	city	total_runs_x	curr_run_rate	req_run_rate	wicke
0	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	0.000000	10.487395	
1	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	0.000000	10.576271	
2	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	8.000000	10.461538	
3	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	6.000000	10.551724	
4	Sunrisers Hyderabad	Royal Challengers Bangalore	Hyderabad	208	7.200000	10.539130	
...	...	...	...	...	...	...	
149572	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.808696	-180.000000	
149573	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.724138	-226.500000	
149574	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.692308	-306.000000	
149575	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.610169	-462.000000	
149576	Chennai Super Kings	Mumbai Indians	Hyderabad	153	15.579832	-936.000000	

147063 rows × 8 columns

In [ ]:

In [54]:
```python
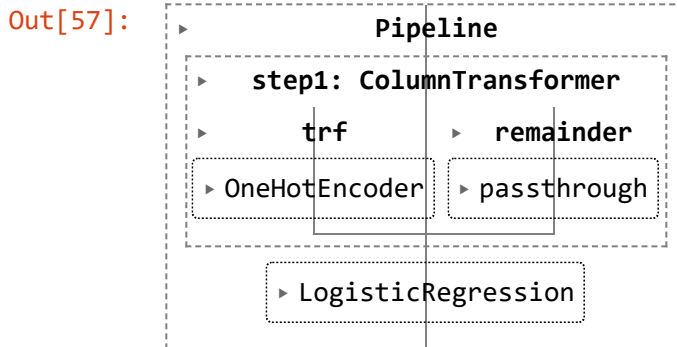# from sklearn.preprocessing import StandardScaler
# sc=StandardScaler()
# x = sc.fit_transform(x)
```

In [55]:
```python
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.2,random_state=1
print('x_train shape',x_train.shape,'y_train shape',y_train.shape)
print('y_test shape',x_test.shape,'y_test shape',y_test.shape)
```

```
x_train shape (117650, 8) y_train shape (117650,)
y_test shape (29413, 8) y_test shape (29413,)
```

```
In [56]:   from sklearn.linear_model import LogisticRegression
           from sklearn.ensemble import RandomForestClassifier,AdaBoostClassifier,Gradier
           from sklearn.tree import DecisionTreeClassifier
           dt=DecisionTreeClassifier()
           lr=LogisticRegression()
           rf=RandomForestClassifier()
           ad=AdaBoostClassifier()
```

```
In [57]:   pipe=Pipeline(steps=[('step1',cf),('step2',LogisticRegression(solver='liblinea
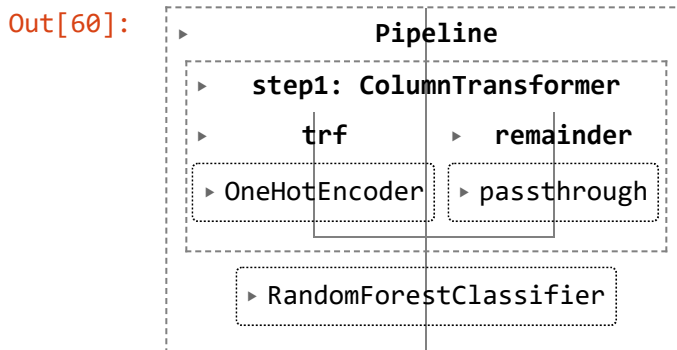           pipe.fit(x_train,y_train)
```

Out[57]:


```
In [58]:   ypred=pipe.predict(x_test)
```

```
In [59]:   pipe.predict_proba(x_test)[10]
```

Out[59]:   array([0.53835726, 0.46164274])

```
In [60]:   pipe1=Pipeline(steps=[('step1',cf),('step2',RandomForestClassifier())])
           pipe1.fit(x_train,y_train)
```

Out[60]:


```
In [61]:   pipe1.predict_proba(x_test)[10]
```

Out[61]:   array([0.97, 0.03])

```python
In [62]: import pickle
         pickle.dump(pipe,open('pipe.pkl','wb'))
```

```
In [ ]:
```