# DDS Project Data Extraction

### November 26, 2023

```
import pandas as pd
     file_path = '/content/all_matches.csv'
     data = pd.read_csv(file_path)
[]: data
[]:
            match_id
                                start_date
                                                                       venue
                                                                              innings
                        season
     0
              682897
                       2013/14
                                2014-03-16
                                             Shere Bangla National Stadium
                                                                                    1
     1
              682897
                       2013/14
                                2014-03-16
                                             Shere Bangla National Stadium
                                                                                    1
     2
              682897
                       2013/14
                                             Shere Bangla National Stadium
                                                                                    1
                                2014-03-16
     3
              682897
                       2013/14
                                2014-03-16
                                             Shere Bangla National Stadium
                                                                                    1
     4
              682897
                       2013/14
                                2014-03-16
                                             Shere Bangla National Stadium
                                                                                    1
     35256
             1298179
                       2022/23
                                2022-11-13
                                                  Melbourne Cricket Ground
                                                                                    2
                                                  Melbourne Cricket Ground
                                                                                    2
     35257
             1298179
                       2022/23
                                2022-11-13
     35258
             1298179
                       2022/23
                                2022-11-13
                                                  Melbourne Cricket Ground
                                                                                    2
                       2022/23
                                                  Melbourne Cricket Ground
                                                                                    2
     35259
             1298179
                                2022-11-13
                                                                                    2
     35260
             1298179
                       2022/23
                                2022-11-13
                                                  Melbourne Cricket Ground
            ball batting_team bowling_team
                                                        striker
                                                                    non_striker
                                              Mohammad Shahzad
     0
             0.1
                   Afghanistan
                                 Bangladesh
                                                                 Najeeb Tarakai
     1
             0.2
                  Afghanistan
                                 Bangladesh
                                                 Gulbadin Naib
                                                                 Najeeb Tarakai
     2
             0.3
                  Afghanistan
                                 Bangladesh
                                                 Gulbadin Naib
                                                                 Najeeb Tarakai
     3
             0.4
                   Afghanistan
                                 Bangladesh
                                                 Gulbadin Naib
                                                                 Najeeb Tarakai
             0.5
                   Afghanistan
                                 Bangladesh
                                                 Gulbadin Naib
                                                                 Najeeb Tarakai
     35256 18.2
                       England
                                   Pakistan
                                                         MM Ali
                                                                      BA Stokes
     35257
            18.3
                       England
                                   Pakistan
                                                LS Livingstone
                                                                      BA Stokes
     35258
            18.4
                       England
                                                     BA Stokes
                                                                 LS Livingstone
                                   Pakistan
                       England
     35259
            18.5
                                   Pakistan
                                                     BA Stokes
                                                                 LS Livingstone
     35260
            18.6
                       England
                                   Pakistan
                                                     BA Stokes
                                                                 LS Livingstone
           extras
                   wides
                           noballs
                                    byes
                                           legbyes
                                                    penalty
                                                              wicket_type
     0
                0
                      NaN
                                      NaN
                                               NaN
                                                                   caught
                               NaN
                                                         NaN
     1
                 1
                      1.0
                                      NaN
                               NaN
                                               NaN
                                                         NaN
                                                                       NaN
     2
                0
                      NaN
                               NaN
                                      NaN
                                               NaN
                                                         NaN
                                                                      NaN
     3
                      NaN
                               NaN
                                      NaN
                                               NaN
                                                         NaN
                                                                      NaN
```

```
NaN
                                                NaN
                                                          NaN
     35256
                 0
                      NaN
                                NaN
                                       NaN
                                                 NaN
                                                          NaN
                                                                     bowled
     35257
                 0
                      NaN
                                NaN
                                       NaN
                                                 NaN
                                                          NaN
                                                                         NaN
     35258
                 0
                      NaN
                                NaN
                                       NaN
                                                 NaN
                                                          NaN
                                                                         NaN
                      NaN
                                                 NaN
                                                                         NaN
     35259
                 0
                                NaN
                                       NaN
                                                          NaN
     35260
                 0
                      {\tt NaN}
                                NaN
                                       NaN
                                                 NaN
                                                          NaN
                                                                         NaN
            player_dismissed other_wicket_type other_player_dismissed
     0
            Mohammad Shahzad
                                              NaN
     1
                           NaN
                                              NaN
                                                                       NaN
     2
                           NaN
                                              NaN
                                                                        NaN
     3
                           NaN
                                              NaN
                                                                       NaN
     4
                           NaN
                                              NaN
                                                                       NaN
     35256
                       MM Ali
                                              NaN
                                                                       NaN
     35257
                           NaN
                                              NaN
                                                                       NaN
     35258
                           NaN
                                              NaN
                                                                       {\tt NaN}
     35259
                           NaN
                                              NaN
                                                                        NaN
     35260
                           {\tt NaN}
                                              NaN
                                                                       NaN
     [35261 rows x 22 columns]
[]: data['venue'].unique()
     data['venue_id'], unique_values = pd.factorize(data['venue'])
[]: data['venue'], data['venue_id']
[]:(0
                Shere Bangla National Stadium
                Shere Bangla National Stadium
      1
      2
                Shere Bangla National Stadium
      3
                Shere Bangla National Stadium
      4
                Shere Bangla National Stadium
      35256
                     Melbourne Cricket Ground
      35257
                     Melbourne Cricket Ground
      35258
                     Melbourne Cricket Ground
      35259
                     Melbourne Cricket Ground
      35260
                     Melbourne Cricket Ground
      Name: venue, Length: 35261, dtype: object,
      1
                 0
      2
                 0
      3
                 0
      4
                 0
      35256
                18
```

4

0

NaN

NaN

NaN

```
35257
               18
      35258
               18
      35259
               18
      35260
               18
     Name: venue_id, Length: 35261, dtype: int64)
[]: batting team players = data.groupby('batting team')['striker', 'non_striker'].
     →agg(lambda x: list(set(x))).reset_index()
     bowling_team_players = data.groupby('bowling_team')['bowler'].agg(lambda x:__
      ⇔list(set(x))).reset_index()
     country = {}
     for index, row in batting_team_players.iterrows():
         if row['batting_team'] in country:
             country[row['batting_team']].append(row['striker'])
         else:
             country[row['batting_team']] = [row['striker']]
     for index, row in bowling_team_players.iterrows():
         if row['bowling_team'] in country:
             country[row['bowling team']].append(row['bowler'])
         else:
             country[row['bowling_team']] = [row['bowler']]
[]: country
[]: unique values dict = {}
     for key, value in country.items():
       unique_values_dict[key] = list(set(value[0] + value[1]))
[]: unique_values_dict
[]: import csv
     with open('players_list.csv', 'w', newline='') as f:
         writer = csv.writer(f)
         writer.writerow(['Key', 'Values'])
         for key, values in unique_values_dict.items():
             for value in values:
                 writer.writerow([key, value])
[]: data[['over_no', 'bowl_no']] = data['ball'].astype(str).str.split('.', __
     ⊶expand=True)
     data['catcher_id'] = None
     data['wicket'] = data['wicket_type'].notna()
     print(data)
           match_id season start_date
                                                                   venue innings \
```

```
0
                            2014-03-16
                                         Shere Bangla National Stadium
         682897
                  2013/14
                                                                                  1
1
         682897
                  2013/14
                            2014-03-16
                                         Shere Bangla National Stadium
                                                                                  1
2
                  2013/14
                            2014-03-16
                                         Shere Bangla National Stadium
                                                                                  1
         682897
3
                                         Shere Bangla National Stadium
         682897
                  2013/14
                            2014-03-16
                                                                                  1
                                         Shere Bangla National Stadium
4
         682897
                  2013/14
                            2014-03-16
                                                                                  1
                                                                                  2
35256
         1298179
                  2022/23
                            2022-11-13
                                               Melbourne Cricket Ground
35257
        1298179
                  2022/23
                            2022-11-13
                                               Melbourne Cricket Ground
                                                                                  2
                                               Melbourne Cricket Ground
                                                                                  2
35258
         1298179
                  2022/23
                            2022-11-13
                                                                                  2
35259
         1298179
                  2022/23
                            2022-11-13
                                               Melbourne Cricket Ground
                  2022/23
                                                                                  2
35260
                            2022-11-13
                                               Melbourne Cricket Ground
         1298179
       ball batting_team bowling_team
                                                     striker
                                                                  non_striker
0
        0.1
              Afghanistan
                             Bangladesh
                                           Mohammad Shahzad
                                                              Najeeb Tarakai
1
        0.2
              Afghanistan
                             Bangladesh
                                              Gulbadin Naib
                                                              Najeeb Tarakai
2
        0.3
              Afghanistan
                             Bangladesh
                                              Gulbadin Naib
                                                              Najeeb Tarakai
3
        0.4
              Afghanistan
                             Bangladesh
                                              Gulbadin Naib
                                                              Najeeb Tarakai
4
        0.5
              Afghanistan
                             Bangladesh
                                              Gulbadin Naib
                                                              Najeeb Tarakai
35256
       18.2
                  England
                                                      MM Ali
                                                                    BA Stokes
                               Pakistan
35257
       18.3
                  England
                               Pakistan
                                             LS Livingstone
                                                                    BA Stokes
35258
       18.4
                  England
                               Pakistan
                                                  BA Stokes
                                                              LS Livingstone
35259
       18.5
                  England
                               Pakistan
                                                  BA Stokes
                                                              LS Livingstone
35260
                  England
                                                              LS Livingstone
       18.6
                               Pakistan
                                                  BA Stokes
                                                            other_wicket_type
      legbyes
                penalty
                          wicket_type
                                        player_dismissed
0
          NaN
                    NaN
                                caught
                                        Mohammad Shahzad
                                                                            NaN
1
          NaN
                    NaN
                                   NaN
                                                       NaN
                                                                            NaN
2
          NaN
                    NaN
                                   NaN
                                                       NaN
                                                                            NaN
3
          NaN
                    NaN
                                   NaN
                                                                            NaN
                                                       NaN
4
          NaN
                                   NaN
                    NaN
                                                       NaN
                                                                            NaN
35256
          NaN
                    NaN
                                bowled
                                                   MM Ali
                                                                            NaN
35257
                                                       NaN
          NaN
                    NaN
                                   NaN
                                                                            NaN
35258
          NaN
                    NaN
                                   NaN
                                                       NaN
                                                                            NaN
35259
          NaN
                    NaN
                                   NaN
                                                       NaN
                                                                            NaN
35260
          NaN
                    NaN
                                   NaN
                                                       NaN
                                                                            NaN
       other_player_dismissed
                                            bowl_no catcher_id wicket
                                  over_no
0
                            NaN
                                        0
                                                  1
                                                           None
                                                                   True
1
                            NaN
                                        0
                                                  2
                                                           None
                                                                 False
2
                                        0
                                                  3
                            NaN
                                                           None
                                                                 False
3
                            NaN
                                        0
                                                  4
                                                                  False
                                                           None
4
                                        0
                                                  5
                            NaN
                                                           None
                                                                  False
35256
                            NaN
                                       18
                                                  2
                                                           None
                                                                   True
35257
                            NaN
                                       18
                                                  3
                                                           None
                                                                  False
35258
                            NaN
                                       18
                                                                  False
                                                           None
```

```
35259 NaN 18 5 None False
35260 NaN 18 6 None False
```

[35261 rows x 26 columns]

```
[]: data2 = pd.read_csv('/content/team.csv')
```

\

## []: data2

[]:		$team\_id$	team_name	coach	home_ground
	0	1	Afghanistan	Graham Reid	1
	1	2	Australia	Andrew McDonald	2
	2	3	${\tt Bangladesh}$	Russell Domingo	3
	3	4	England	Matthew Mott	4
	4	5	Hong Kong	Simon Muggleton	5
	5	6	India	Rahul Dravid	6
	6	7	Ireland	Graham Ford	7
	7	8	Namibia	Pierre de Bruyn	8
	8	9	Nepal	Pubudu Dassanayake	9
	9	10	Netherlands	Ryan Campbell	10
	10	11	New Zealand	Gary Stead	11
	11	12	Oman	Duleep Mendis	12
	12	13	Pakistan	Saqlain Mushtaq	13
	13	14	Papua New Guinea	Simon Lavington	14
	14	15	Scotland	Shane Burger	15
	15	16	South Africa	Mark Boucher	16
	16	17	Sri Lanka	Chris Silverwood	17
	17	18	United Arab Emirates	Robin Singh	18
	18	19	West Indies	Phil Simmons	19
	19	20	Zimbabwe	Dave Houghton	20

owner

0 Afghanistan Cricket Board Cricket Australia 1 2 Bangladesh Cricket Board 3 England and Wales Cricket Board 4 Hong Kong Cricket Association 5 Board of Control for Cricket in India 6 Cricket Ireland 7 Cricket Namibia 8 Cricket Association of Nepal Koninklijke Nederlandse Cricket Bond 9 10 New Zealand Cricket Oman Cricket 11 12 Pakistan Cricket Board 13 Cricket Papua New Guinea 14 Cricket Scotland

```
15
                           Cricket South Africa
     16
                              Sri Lanka Cricket
     17
                         Emirates Cricket Board
     18
                            Cricket West Indies
     19
                               Zimbabwe Cricket
[]: data2 = data2[['team_id','team_name']]
     data2 = data2.rename(columns={'team_id':'batting_team_id'})
[]:
     data2
[]:
         batting_team_id
                                      team_name
     0
                        1
                                    Afghanistan
     1
                        2
                                      Australia
     2
                        3
                                     Bangladesh
                        4
     3
                                        England
     4
                        5
                                      Hong Kong
     5
                        6
                                          India
     6
                       7
                                        Treland
     7
                       8
                                        Namibia
     8
                       9
                                          Nepal
     9
                       10
                                    Netherlands
                                    New Zealand
     10
                       11
     11
                       12
                                            Oman
     12
                       13
                                       Pakistan
                               Papua New Guinea
     13
                       14
     14
                                       Scotland
                       15
     15
                       16
                                   South Africa
                       17
                                      Sri Lanka
     16
     17
                       18
                           United Arab Emirates
                                    West Indies
     18
                       19
     19
                       20
                                       Zimbabwe
[]: data = data.reset_index().merge(data2, how='left', left_on='batting_team',_
      oright_on='team_name').set_index('index')
     print(data)
           match_id
                       season
                               start_date
                                                                     venue
                                                                            innings \
    index
    0
             682897
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
                                                                                   1
    1
             682897
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
    2
              682897
                      2013/14 2014-03-16
                                            Shere Bangla National Stadium
                                                                                   1
                                            Shere Bangla National Stadium
    3
              682897
                      2013/14
                               2014-03-16
                                                                                   1
             682897
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
                                                                                   1
             1298179 2022/23 2022-11-13
                                                                                   2
    35256
                                                 Melbourne Cricket Ground
```

35257 35258 35259 35260	1298 1298 1298 1298	3179 3179	2022/23 2022/23 2022/23 2022/23	2022 2022	-11-13 -11-13 -11-13 -11-13	Melbo Melbo	urne Cri urne Cri urne Cri urne Cri	cket cket	Ground Ground			2 2 2 2
index	ball	batt:	ing_team	bowli	ng_team	:	striker		non_str	iker		\
0	0.1	Afgl	hanistan	Ban	gladesh	Mohammad S	Shahzad	Naj	eeb Tara	akai	•••	
1	0.2 Afghanistan		Ban	gladesh	Gulbadin Naib		Naj	eeb Tara	akai	•••		
2	0.3 Afghanistan			Ban	gladesh	Gulbadin Naib Najeeb			eeb Tara	akai	•••	
3	0.4 Afghanistan			Ban	gladesh	Gulbadin Naib Najeeb		eeb Tara	akai			
4	0.5 Afghanistan			Ban	gladesh	Gulbadin Naib Najeeb Tara		akai				
•••	•••		•••	•••		•••			•••			
35256	18.2		England	P	akistan		MM Ali		BA Sto	okes		
35257	18.3		England	P	akistan	LS Livi	ngstone		BA Sto	okes		
35258	18.4		England	P	akistan		Stokes	LS	Livingst	tone		
35259	18.5		England	Р	akistan	ВА	Stokes		Livings		•••	
35260	18.6		England	Р	akistan	ВА	Stokes		Livings		•••	
<pre>wicket_type player_dismissed other_wicket_type \</pre>												
index		_ 71	1 3	_		_	_ J1					
0	c	aught	t Mohamn	nad Sh	ahzad		NaN					
1	NaN				NaN							
2	NaN				NaN NaN							
3		Nal			NaN NaN							
4	NaN NaN				NaN							
		IVai	·		wan		wan					
 35256	1·	 oowled	4	м	M Ali	•••	NaN					
35257	NaN			NaN		NaN						
35258		Nal			NaN NaN							
35259	NaN NaN				NaN NaN							
35260	NaN				NaN NaN							
55200		IVai	·V		IVAIV		wan					
	other_player_dismissed over_no bowl_no catcher_id wicket \									\		
index												
0				NaN	0	1	N	one	True			
1				NaN	0	2	N	one	False			
2				NaN	0	3	N	one	False			
3				NaN	0	4	N	one	False			
4				NaN	0	5	N	one	False			
			••	•								
35256				NaN	18	2	N	one	True			
35257				NaN	18	3	N	one	False			
35258				NaN	18	4	N	one	False			
35259				NaN	18	5	N	one	False			
35260				NaN	18	6	N	one	False			

batting\_team\_id team\_name

```
0
                             Afghanistan
    1
                          1
                             Afghanistan
    2
                             Afghanistan
    3
                             Afghanistan
    4
                             Afghanistan
    35256
                          4
                                 England
    35257
                          4
                                 England
    35258
                          4
                                 England
    35259
                          4
                                 England
    35260
                          4
                                 England
    [35261 rows x 28 columns]
[]: data2 = data2.rename(columns={'batting_team_id':'bowling_team_id'})
[]:
    data2
[]:
         bowling_team_id
                                       team_name
     0
                                     Afghanistan
     1
                        2
                                       Australia
     2
                        3
                                      Bangladesh
     3
                        4
                                         England
     4
                        5
                                       Hong Kong
                        6
     5
                                            India
     6
                        7
                                         Ireland
     7
                        8
                                         Namibia
                        9
     8
                                            Nepal
     9
                       10
                                     Netherlands
                                     New Zealand
     10
                       11
                                             Oman
     11
                       12
     12
                       13
                                        Pakistan
     13
                       14
                                Papua New Guinea
                                        Scotland
     14
                       15
     15
                       16
                                    South Africa
     16
                       17
                                       Sri Lanka
     17
                       18
                           United Arab Emirates
                       19
                                     West Indies
     18
     19
                       20
                                        Zimbabwe
[]: data = data.reset_index().merge(data2, how='left', left_on='bowling_team',_
      →right_on='team_name').set_index('index')
     print(data)
           {\tt match\_id}
                                start_date
                                                                              innings \
                       season
                                                                       venue
    index
    0
                      2013/14 2014-03-16 Shere Bangla National Stadium
              682897
                                                                                     1
```

index

```
Shere Bangla National Stadium
1
         682897
                  2013/14
                           2014-03-16
                                                                                1
2
         682897
                  2013/14
                            2014-03-16
                                        Shere Bangla National Stadium
                                                                                 1
3
                  2013/14
                            2014-03-16
                                        Shere Bangla National Stadium
                                                                                1
         682897
4
                  2013/14
                                        Shere Bangla National Stadium
         682897
                            2014-03-16
                                                                                 1
                                              Melbourne Cricket Ground
                                                                                2
35256
        1298179
                  2022/23
                            2022-11-13
35257
        1298179
                  2022/23
                            2022-11-13
                                              Melbourne Cricket Ground
                                                                                2
35258
        1298179
                  2022/23
                            2022-11-13
                                              Melbourne Cricket Ground
                                                                                2
                                              Melbourne Cricket Ground
                                                                                2
35259
        1298179
                  2022/23
                            2022-11-13
35260
        1298179
                  2022/23
                           2022-11-13
                                              Melbourne Cricket Ground
                                                                                2
       ball batting_team bowling_team
                                                   striker
                                                                non_striker
index
0
        0.1
             Afghanistan
                             Bangladesh
                                          Mohammad Shahzad
                                                             Najeeb Tarakai
1
        0.2
              Afghanistan
                             Bangladesh
                                             Gulbadin Naib
                                                             Najeeb Tarakai
2
                                             Gulbadin Naib
        0.3
             Afghanistan
                             Bangladesh
                                                             Najeeb Tarakai
3
        0.4
             Afghanistan
                             Bangladesh
                                             Gulbadin Naib
                                                             Najeeb Tarakai
4
        0.5
             Afghanistan
                             Bangladesh
                                             Gulbadin Naib
                                                             Najeeb Tarakai
35256
       18.2
                  England
                                                    MM Ali
                                                                  BA Stokes
                               Pakistan
35257
       18.3
                  England
                               Pakistan
                                            LS Livingstone
                                                                  BA Stokes
35258
       18.4
                  England
                               Pakistan
                                                 BA Stokes
                                                             LS Livingstone
35259
       18.5
                  England
                               Pakistan
                                                 BA Stokes
                                                             LS Livingstone
                                                 BA Stokes
                                                             LS Livingstone
35260
       18.6
                  England
                               Pakistan
                          other_player_dismissed
      other_wicket_type
                                                   over_no
                                                              bowl_no
                                                                       catcher_id
index
0
                     NaN
                                               NaN
                                                           0
                                                                     1
                                                                              None
                                                           0
                                                                     2
1
                     NaN
                                               NaN
                                                                              None
2
                     NaN
                                               NaN
                                                           0
                                                                     3
                                                                              None
3
                                                           0
                                                                     4
                     NaN
                                               NaN
                                                                              None
4
                     NaN
                                               NaN
                                                           0
                                                                     5
                                                                              None
                                                                     ...
                                                                     2
35256
                     NaN
                                               {\tt NaN}
                                                          18
                                                                              None
35257
                                               NaN
                                                                     3
                                                                              None
                     NaN
                                                          18
35258
                     NaN
                                               NaN
                                                          18
                                                                     4
                                                                              None
                                                                     5
35259
                     NaN
                                               NaN
                                                          18
                                                                              None
35260
                     NaN
                                               NaN
                                                          18
                                                                     6
                                                                              None
               batting_team_id team_name_x bowling_team_id team_name_y
       wicket
index
0
         True
                                  Afghanistan
                                                                 Bangladesh
1
        False
                                  Afghanistan
                                                              3
                                                                 Bangladesh
2
                                                              3
        False
                                  Afghanistan
                                                                 Bangladesh
3
        False
                                  Afghanistan
                                                              3
                                                                 Bangladesh
                               1
4
        False
                               1
                                  Afghanistan
                                                              3
                                                                 Bangladesh
35256
                               4
                                      England
                                                             13
                                                                   Pakistan
         True
```

```
35257
            False
                                  4
                                          England
                                                                13
                                                                      Pakistan
    35258
            False
                                  4
                                          England
                                                                13
                                                                      Pakistan
    35259
            False
                                  4
                                          England
                                                                13
                                                                      Pakistan
    35260
            False
                                          England
                                                                13
                                                                      Pakistan
    [35261 rows x 30 columns]
[]: data3 = pd.read_csv('/content/players.csv')
     data3 = data3[['player_id','player_name','team_id']]
     data3 = data3.rename(columns={'player_id':'batsman_id'})
[]:
     data3
[]:
          batsman_id
                              player_name
                                           team_id
     0
                               Aftab Alam
                   1
                                                  1
     1
                   2
                               Amir Hamza
                                                  1
                   3
     2
                         Asghar Stanikzai
                                                  1
     3
                   4
                      Azmatullah Omarzai
                                                  1
     4
                   5
                          Darwish Rasooli
                                                  1
     505
                 506
                             T Panyangara
                                                 20
     506
                 507
                               TL Chatara
                                                 20
     507
                 508
                                V Sibanda
                                                 20
                              W Madhevere
     508
                 509
                                                 20
     509
                 510
                             WP Masakadza
                                                 20
     [510 rows x 3 columns]
[]: data = data.reset_index().merge(data3, how='left',__
      ⇔left_on=['striker','batting_team_id'], right_on=['player_name','team_id']).
      ⇔set_index('index')
     print(data)
                                                                           innings \
           {\tt match\_id}
                       season start_date
                                                                     venue
    index
    0
                      2013/14
                                            Shere Bangla National Stadium
                                                                                   1
             682897
                               2014-03-16
    1
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
             682897
                                                                                   1
    2
             682897
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
                                                                                   1
    3
             682897
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
                                                                                   1
    4
                                            Shere Bangla National Stadium
             682897
                      2013/14
                               2014-03-16
                                                                                   1
                      2022/23
                                                 Melbourne Cricket Ground
                                                                                   2
    35256
             1298179
                               2022-11-13
    35257
             1298179
                     2022/23
                               2022-11-13
                                                 Melbourne Cricket Ground
                                                                                   2
                                                 Melbourne Cricket Ground
                                                                                   2
    35258
             1298179
                      2022/23
                               2022-11-13
                                                                                   2
    35259
                      2022/23
                                                 Melbourne Cricket Ground
             1298179
                               2022-11-13
    35260
             1298179
                     2022/23 2022-11-13
                                                 Melbourne Cricket Ground
                                                                                   2
```

```
ball batting_team bowling_team
                                                    striker
                                                                non_striker
index
0
        0.1
              Afghanistan
                             Bangladesh
                                          Mohammad Shahzad
                                                             Najeeb Tarakai
1
        0.2
              Afghanistan
                             Bangladesh
                                                             Najeeb Tarakai
                                             Gulbadin Naib
2
        0.3
              Afghanistan
                             Bangladesh
                                             Gulbadin Naib
                                                             Najeeb Tarakai
3
        0.4
              Afghanistan
                             Bangladesh
                                             Gulbadin Naib
                                                             Najeeb Tarakai
4
        0.5
              Afghanistan
                             Bangladesh
                                             Gulbadin Naib
                                                             Najeeb Tarakai
35256
      18.2
                  England
                               Pakistan
                                                    MM Ali
                                                                   BA Stokes
35257
       18.3
                  England
                                            LS Livingstone
                                                                   BA Stokes
                               Pakistan
                  England
35258
       18.4
                               Pakistan
                                                 BA Stokes
                                                             LS Livingstone
35259
                  England
                                                             LS Livingstone
       18.5
                               Pakistan
                                                 BA Stokes
                  England
35260
       18.6
                               Pakistan
                                                 BA Stokes
                                                             LS Livingstone
                catcher_id wicket
                                     batting_team_id team_name_x
      bowl_no
index
0
             1
                      None
                               True
                                                        Afghanistan
1
             2
                      None
                              False
                                                        Afghanistan
                                                     1
2
             3
                      None
                              False
                                                        Afghanistan
                                                        Afghanistan
3
             4
                      None
                              False
4
                                                        Afghanistan
             5
                      None
                              False
35256
             2
                               True
                                                            England
                      None
                                                     4
35257
             3
                      None
                              False
                                                     4
                                                            England
             4
                                                     4
                                                            England
35258
                      None
                              False
35259
             5
                                                     4
                                                            England
                      None
                              False
                                                     4
35260
                      None
                              False
                                                            England
       bowling_team_id team_name_y
                                        batsman_id
                                                          player_name team_id
index
0
                      3
                           Bangladesh
                                                17
                                                     Mohammad Shahzad
                                                                              1
                      3
                                                 9
1
                           Bangladesh
                                                        Gulbadin Naib
                                                                              1
2
                      3
                          Bangladesh
                                                 9
                                                        Gulbadin Naib
                                                                              1
3
                      3
                           Bangladesh
                                                 9
                                                        Gulbadin Naib
                                                                              1
4
                      3
                           Bangladesh
                                                 9
                                                        Gulbadin Naib
                                                                              1
35256
                     13
                             Pakistan
                                               119
                                                               MM Ali
                                                                              4
35257
                                               116
                                                       LS Livingstone
                                                                              4
                     13
                             Pakistan
                                                            BA Stokes
35258
                     13
                             Pakistan
                                               101
                                                                              4
35259
                     13
                             Pakistan
                                               101
                                                            BA Stokes
                                                                              4
35260
                                                                              4
                     13
                             Pakistan
                                               101
                                                            BA Stokes
[35261 rows x 33 columns]
```

[]: data3 = data3.rename(columns={'batsman\_id':'non\_striker\_id'})

#### []: data3 []: non\_striker\_id player\_name team id Aftab Alam 2 1 Amir Hamza 1 2 3 Asghar Stanikzai 1 3 4 Azmatullah Omarzai 1 4 5 Darwish Rasooli 1 505 506 T Panyangara 20 20 506 507 TL Chatara 507 508 V Sibanda 20 20 508 509 W Madhevere 509 510 WP Masakadza 20 [510 rows x 3 columns] []: data = data.reset\_index().merge(data3, how='left',\_ ⇔left\_on=['non\_striker','batting\_team\_id'], →right\_on=['player\_name','team\_id']).set\_index('index') print(data) ${\tt match\_id}$ innings season start\_date venue index 0 682897 2013/14 2014-03-16 Shere Bangla National Stadium 1 1 682897 2013/14 2014-03-16 Shere Bangla National Stadium 1 2 1 682897 2013/14 2014-03-16 Shere Bangla National Stadium 3 2013/14 Shere Bangla National Stadium 682897 2014-03-16 1 4 2013/14 2014-03-16 Shere Bangla National Stadium 682897 1 1298179 2022/23 2022-11-13 Melbourne Cricket Ground 2 35256 35257 1298179 2022/23 2022-11-13 Melbourne Cricket Ground 2 35258 1298179 2022/23 2022-11-13 Melbourne Cricket Ground 2 35259 2022/23 2022-11-13 Melbourne Cricket Ground 2 1298179 Melbourne Cricket Ground 2 35260 1298179 2022/23 2022-11-13 ball batting\_team bowling\_team striker non\_striker index 0 0.1 Mohammad Shahzad Afghanistan Bangladesh Najeeb Tarakai 1 0.2 Afghanistan Bangladesh Gulbadin Naib Najeeb Tarakai 2 0.3 Afghanistan Bangladesh Gulbadin Naib Najeeb Tarakai 3 0.4 Afghanistan Bangladesh Gulbadin Naib Najeeb Tarakai 4 0.5 Afghanistan Najeeb Tarakai Bangladesh Gulbadin Naib 35256 18.2 England Pakistan MM Ali BA Stokes 35257 England Pakistan BA Stokes 18.3 LS Livingstone 35258 18.4 England Pakistan BA Stokes LS Livingstone

```
35259 18.5
                      England
                                                   BA Stokes LS Livingstone
                                  Pakistan
    35260 18.6
                      England
                                  Pakistan
                                                   BA Stokes
                                                              LS Livingstone
          batting_team_id team_name_x bowling_team_id team_name_y
                                                                        batsman_id \
    index
    0
                            Afghanistan
                                                            Bangladesh
                                                                                 17
    1
                           Afghanistan
                                                        3
                                                            Bangladesh
                                                                                 9
    2
                         1
                           Afghanistan
                                                            Bangladesh
                                                                                  9
    3
                         1 Afghanistan
                                                        3
                                                            Bangladesh
                                                                                  9
    4
                           Afghanistan
                                                        3
                                                            Bangladesh
                                                                                  9
                         4
    35256
                                England
                                                       13
                                                              Pakistan
                                                                                119
    35257
                                England
                         4
                                                              Pakistan
                                                                                116
                                                       13
                         4
    35258
                                England
                                                       13
                                                              Pakistan
                                                                                101
                         4
    35259
                                England
                                                       13
                                                              Pakistan
                                                                                101
    35260
                                England
                                                       13
                                                              Pakistan
                                                                                101
              player_name_x team_id_x non_striker_id
                                                           player_name_y team_id_y
    index
    0
           Mohammad Shahzad
                                      1
                                                      19 Najeeb Tarakai
                                                                                  1
    1
              Gulbadin Naib
                                                          Najeeb Tarakai
                                      1
                                                                                  1
    2
              Gulbadin Naib
                                      1
                                                      19
                                                          Najeeb Tarakai
                                                                                  1
    3
              Gulbadin Naib
                                      1
                                                          Najeeb Tarakai
                                                                                  1
    4
              Gulbadin Naib
                                                          Najeeb Tarakai
                                      1
                                                      19
                                                                                  1
                       •••
                                      4
                                                     101
                                                                                  4
    35256
                     MM Ali
                                                               BA Stokes
    35257
                                      4
                                                     101
                                                               BA Stokes
                                                                                  4
             LS Livingstone
                                      4
                                                                                  4
    35258
                  BA Stokes
                                                     116 LS Livingstone
    35259
                  BA Stokes
                                      4
                                                                                  4
                                                     116
                                                          LS Livingstone
    35260
                  BA Stokes
                                      4
                                                     116
                                                          LS Livingstone
    [35261 rows x 36 columns]
[]: data3 = data3.rename(columns={'non_striker_id':'bowler_id'})
[]: data = data.reset_index().merge(data3, how='left',__
      →left_on=['bowler','bowling_team_id'], right_on=['player_name','team_id']).
      ⇔set_index('index')
     print(data)
           match_id
                               start_date
                                                                           innings
                       season
                                                                    venue
    index
    0
             682897
                     2013/14
                               2014-03-16
                                           Shere Bangla National Stadium
                                                                                  1
    1
             682897 2013/14 2014-03-16 Shere Bangla National Stadium
                                                                                  1
    2
             682897 2013/14 2014-03-16 Shere Bangla National Stadium
                                                                                  1
    3
             682897 2013/14 2014-03-16
                                           Shere Bangla National Stadium
                                                                                  1
                     2013/14 2014-03-16
    4
             682897
                                           Shere Bangla National Stadium
```

```
35256
        1298179
                 2022/23
                           2022-11-13
                                             Melbourne Cricket Ground
                                                                               2
35257
                                                                               2
        1298179
                 2022/23
                           2022-11-13
                                             Melbourne Cricket Ground
                                                                               2
35258
        1298179
                 2022/23
                           2022-11-13
                                             Melbourne Cricket Ground
35259
                 2022/23
                           2022-11-13
                                             Melbourne Cricket Ground
                                                                               2
        1298179
                                                                               2
                                             Melbourne Cricket Ground
35260
        1298179
                 2022/23
                           2022-11-13
       ball batting_team bowling_team
                                                  striker
                                                               non striker
index
0
             Afghanistan
                            Bangladesh
                                         Mohammad Shahzad
                                                            Najeeb Tarakai
        0.1
1
        0.2
             Afghanistan
                            Bangladesh
                                            Gulbadin Naib
                                                            Najeeb Tarakai
2
        0.3
             Afghanistan
                            Bangladesh
                                            Gulbadin Naib
                                                            Najeeb Tarakai
3
        0.4
             Afghanistan
                            Bangladesh
                                            Gulbadin Naib
                                                            Najeeb Tarakai
4
        0.5
             Afghanistan
                            Bangladesh
                                            Gulbadin Naib
                                                            Najeeb Tarakai
35256
      18.2
                  England
                              Pakistan
                                                   MM Ali
                                                                 BA Stokes
                              Pakistan
                  England
                                                                 BA Stokes
35257
       18.3
                                           LS Livingstone
35258
       18.4
                  England
                              Pakistan
                                                BA Stokes
                                                            LS Livingstone
35259
       18.5
                  England
                              Pakistan
                                                BA Stokes
                                                            LS Livingstone
35260
      18.6
                  England
                              Pakistan
                                                BA Stokes
                                                            LS Livingstone
                                                               non_striker_id
      team name y
                   batsman_id
                                   player_name_x
                                                   team_id_x
index
0
       Bangladesh
                            17
                                Mohammad Shahzad
                                                            1
                                                                            19
1
       Bangladesh
                             9
                                    Gulbadin Naib
                                                            1
                                                                            19
2
       Bangladesh
                             9
                                   Gulbadin Naib
                                                            1
                                                                            19
3
                             9
                                                            1
       Bangladesh
                                   Gulbadin Naib
                                                                            19
4
                             9
                                                            1
       Bangladesh
                                    Gulbadin Naib
                                                                            19
            •••
35256
         Pakistan
                           119
                                           MM Ali
                                                            4
                                                                           101
35257
         Pakistan
                           116
                                  LS Livingstone
                                                            4
                                                                           101
35258
         Pakistan
                           101
                                        BA Stokes
                                                            4
                                                                           116
35259
         Pakistan
                           101
                                        BA Stokes
                                                            4
                                                                           116
35260
         Pakistan
                           101
                                        BA Stokes
                                                                           116
        player_name_y
                       team id y
                                   bowler id
                                                    player_name team_id
index
0
                                                                        3
       Najeeb Tarakai
                                1
                                           72
                                               Mashrafe Mortaza
1
       Najeeb Tarakai
                                1
                                               Mashrafe Mortaza
                                                                        3
2
       Najeeb Tarakai
                                1
                                           72
                                               Mashrafe Mortaza
                                                                        3
3
                                               Mashrafe Mortaza
       Najeeb Tarakai
                                1
                                           72
                                                                        3
4
       Najeeb Tarakai
                                1
                                           72
                                               Mashrafe Mortaza
                                                                        3
35256
            BA Stokes
                                4
                                          311
                                                 Mohammad Wasim
                                                                      13
                                4
                                                 Mohammad Wasim
                                                                      13
35257
            BA Stokes
                                          311
35258
       LS Livingstone
                                4
                                          311
                                                 Mohammad Wasim
                                                                      13
35259
       LS Livingstone
                                          311
                                                 Mohammad Wasim
                                                                      13
35260
       LS Livingstone
                                          311
                                                 Mohammad Wasim
                                                                      13
```

```
[]:|data3 = data3.rename(columns={'bowler_id':'player_out_id'})
[]: data = data.reset_index().merge(data3, how='left',_
      ⇔left_on=['player_dismissed','batting_team_id'],
      →right_on=['player_name','team_id']).set_index('index')
     print(data)
           match_id
                       season
                               start_date
                                                                     venue
                                                                             innings
    index
    0
              682897
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
                                                                                   1
    1
              682897
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
                                                                                   1
    2
              682897
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
                                                                                   1
    3
              682897
                      2013/14
                               2014-03-16
                                            Shere Bangla National Stadium
                                                                                   1
    4
                                            Shere Bangla National Stadium
              682897
                      2013/14
                               2014-03-16
                                                                                   1
                                                                                   2
                                                 Melbourne Cricket Ground
    35256
             1298179
                      2022/23
                               2022-11-13
    35257
             1298179
                      2022/23
                               2022-11-13
                                                 Melbourne Cricket Ground
                                                                                   2
                                                                                   2
    35258
             1298179
                      2022/23
                               2022-11-13
                                                 Melbourne Cricket Ground
    35259
                      2022/23
                               2022-11-13
                                                 Melbourne Cricket Ground
                                                                                   2
             1298179
                      2022/23
                                                 Melbourne Cricket Ground
                                                                                   2
    35260
             1298179
                               2022-11-13
           ball batting_team bowling_team
                                                       striker
                                                                   non_striker
    index
    0
            0.1
                  Afghanistan
                                 Bangladesh
                                             Mohammad Shahzad
                                                                Najeeb Tarakai
    1
            0.2
                  Afghanistan
                                Bangladesh
                                                Gulbadin Naib
                                                                Najeeb Tarakai
    2
             0.3
                  Afghanistan
                                Bangladesh
                                                Gulbadin Naib
                                                                Najeeb Tarakai
    3
             0.4
                  Afghanistan
                                Bangladesh
                                                Gulbadin Naib
                                                                Najeeb Tarakai
    4
             0.5
                  Afghanistan
                                Bangladesh
                                                Gulbadin Naib
                                                                Najeeb Tarakai
    35256
          18.2
                      England
                                  Pakistan
                                                       MM Ali
                                                                     BA Stokes
    35257
           18.3
                      England
                                   Pakistan
                                               LS Livingstone
                                                                     BA Stokes
    35258
           18.4
                      England
                                   Pakistan
                                                    BA Stokes
                                                                LS Livingstone
    35259
           18.5
                      England
                                   Pakistan
                                                    BA Stokes
                                                                LS Livingstone
    35260
           18.6
                      England
                                                                LS Livingstone
                                   Pakistan
                                                    BA Stokes
           team_id_x non_striker_id
                                         player_name_y team_id_y
                                                                   bowler_id
    index
    0
                    1
                                    19
                                        Najeeb Tarakai
                                                                          72
    1
                    1
                                        Najeeb Tarakai
                                                                1
                                                                          72
    2
                    1
                                    19
                                        Najeeb Tarakai
                                                                1
                                                                          72
    3
                    1
                                    19
                                        Najeeb Tarakai
                                                                1
                                                                          72
    4
                    1
                                        Najeeb Tarakai
                                                                1
                                                                          72
                                    19
    35256
                    4
                                   101
                                             BA Stokes
                                                                4
                                                                          311
                                             BA Stokes
                    4
                                   101
                                                                4
                                                                          311
    35257
    35258
                                   116 LS Livingstone
                                                                          311
```

```
35260
                                116 LS Livingstone
                                                                   311
                  4
                                                           4
             player_name_x team_id_x player_out_id
                                                        player_name_y team_id_y
    index
    0
          Mashrafe Mortaza
                                   3
                                               17.0 Mohammad Shahzad
                                                                           1.0
    1
          Mashrafe Mortaza
                                   3
                                                NaN
                                                                           NaN
    2
          Mashrafe Mortaza
                                   3
                                                NaN
                                                                 NaN
                                                                           NaN
    3
          Mashrafe Mortaza
                                   3
                                                NaN
                                                                 NaN
                                                                           NaN
          Mashrafe Mortaza
    4
                                   3
                                                NaN
                                                                 NaN
                                                                           NaN
    35256
                                              119.0
                                                              MM Ali
                                                                           4.0
            Mohammad Wasim
                                   13
    35257
            Mohammad Wasim
                                   13
                                                NaN
                                                                 NaN
                                                                           NaN
    35258
            Mohammad Wasim
                                   13
                                                NaN
                                                                 NaN
                                                                           NaN
    35259
            Mohammad Wasim
                                   13
                                                NaN
                                                                 NaN
                                                                           NaN
    35260
            Mohammad Wasim
                                   13
                                                NaN
                                                                 NaN
                                                                           NaN
    [35261 rows x 42 columns]
    <ipython-input-50-8d60fc0a8f09>:1: FutureWarning: Passing 'suffixes' which cause
    duplicate columns {'team_id_x', 'player_name_x'} in the result is deprecated and
    will raise a MergeError in a future version.
      data = data.reset index().merge(data3, how='left',
    left_on=['player_dismissed','batting_team_id'],
    right_on=['player_name','team_id']).set_index('index')
[]: data = data.rename(columns={'innings': 'inning id', 'runs_off_bat':__
     data = data[['match_id', 'inning_id', 'bowl_no', 'over_no', 'wicket',

¬'non_striker_id', 'wicket_type', 'catcher_id', 'player_out_id']]

[]: data.to_csv('new_file.csv', index=False)
[]: len(data['match_id'].unique())
[]: 154
[]: unique batsman ids = data.groupby('match id')['batsman id'].unique()
[]: unique_bowler_ids = data.groupby('match_id')['bowler_id'].unique()
[]: unique non_batsman_ids = data.groupby('match_id')['non_striker_id'].unique()
[]: combined ids = unique batsman ids.combine(unique bowler ids, lambda x1, x2:
     ⇒list(set(list(x1) + list(x2))))
    combined ids2 = combined ids.combine(unique non_batsman_ids, lambda x1, x2:__
      \hookrightarrowlist(set(list(x1) + list(x2))))
```

116 LS Livingstone

4

311

35259

4

#### print(combined\_ids2) match\_id 682897 [1, 6, 9, 14, 16, 17, 19, 22, 26, 27, 28, 60, ... [128, 130, 132, 133, 134, 136, 138, 139, 140, ... 682899 [172, 173, 174, 179, 181, 182, 186, 188, 192, ... 682901 682903 [430, 431, 432, 437, 438, 440, 442, 444, 445, ... 682905 [128, 130, 3, 132, 133, 134, 2, 136, 137, 9, 6... [293, 297, 299, 306, 308, 309, 311, 312, 315, ... 1298175 [147, 148, 149, 152, 156, 158, 160, 163, 164, ... 1298176 [257, 258, 260, 262, 266, 268, 269, 293, 297, ... 1298177 1298178 [147, 148, 149, 152, 156, 158, 160, 163, 164, ... [293, 297, 299, 306, 308, 309, 311, 312, 315, ... 1298179 Length: 154, dtype: object []: df = combined\_ids2.apply(pd.Series).stack().reset\_index(level=1, drop=True). →reset\_index() df.columns = ['Key', 'Value'] df.to\_csv('output2.csv', index=False)