In [1]:

```python
import numpy as np
import pandas as pd
import os
from yaml import safe_load
from tqdm import tqdm
```

In [2]:

```python
filenames = []
for file in os.listdir('data'):
    filenames.append(os.path.join('data', file))
```

In [3]:

```python
filenames[:10]
```

Out[3]:

```
['data\\1001349.yaml',
 'data\\1001351.yaml',
 'data\\1001353.yaml',
 'data\\1004729.yaml',
 'data\\1007655.yaml',
 'data\\1007657.yaml',
 'data\\1007659.yaml',
 'data\\1019979.yaml',
 'data\\1019981.yaml',
 'data\\1019983.yaml']
```

In [4]:

```python
final_df = pd.DataFrame()
counter =1
for file in tqdm(filenames):
    with open(file, 'r') as f:
        df = pd.json_normalize(safe_load(f))
        df['match_id']=counter
        final_df = final_df.append(df)
        counter+=1
final_df
```

```
100%|████████████████████████████████████████████████████████████████████████████| 1432/1433 [04:25<00:00,  5.39it/s]
---------------------------------------------------------------------------
NotImplementedError                       Traceback (most recent call last)
<ipython-input-4-23b969b392fc> in <module>
      3 for file in tqdm(filenames):
      4     with open(file, 'r') as f:
----> 5         df = pd.json_normalize(safe_load(f))
      6         df['match_id']=counter
      7         final_df = final_df.append(df)

c:\users\92336\appdata\local\programs\python\python39\lib\site-packages\pandas\io\json\_normalize.py in _json_normalize(data, record_path, meta, meta_prefix, record_prefix, errors, sep, max_level)
    421         data = list(data)
    422     else:
--> 423         raise NotImplementedError
    424 
    425     # check to see if a simple recursive function is possible to

NotImplementedError:
```

In [5]:

```python
backup = final_df.copy
```

In [6]:

```python
final_df
```

Out[6]:

	innings	meta.data_version	meta.created	meta.revision	info.dates	info.gender	info.match_type	info.
0	[{'1st innings': {'team': 'Australia', 'delive...	0.9	2017-02-18	2	[2017-02-17]	male	T20	
0	[{'1st innings': {'team': 'Australia', 'delive...	0.9	2017-02-19	2	[2017-02-19]	male	T20	
0	[{'1st innings': {'team': 'Australia', 'delive...	0.9	2017-02-23	1	[2017-02-22]	male	T20	
0	[{'1st innings': {'team': 'Hong Kong', 'delive...	0.9	2016-09-12	1	[2016-09-05]	male	T20	
0	[{'1st innings': {'team': 'Zimbabwe', 'deliver...	0.9	2016-06-19	1	[2016-06-18]	male	T20	
...	...	...	...	...	...	...	...	
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	0.9	2016-03-05	2	[2016-03-04]	male	T20	
0	[{'1st innings': {'team': 'Bangladesh', 'deliv...	0.9	2016-03-08	1	[2016-03-06]	male	T20	
0	[{'1st innings': {'team': 'Netherlands', 'deli...	0.9	2016-02-03	1	[2016-02-03]	male	T20	
0	[{'1st innings': {'team': 'Australia', 'delive...	0.9	2016-09-12	1	[2016-09-06]	male	T20	
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	0.9	2016-09-12	1	[2016-09-09]	male	T20	

1432 rows × 28 columns

In [7]:

```python
final_df.drop(columns=[
    'meta.data_version',
    'meta.created',
    'meta.revision',
    'info.outcome.bowl_out',
    'info.bowl_out',
    'info.supersubs.South Africa',
    'info.supersubs.New Zealand',
    'info.outcome.eliminator',
    'info.outcome.result',
    'info.outcome.method',
    'info.neutral_venue',
    'info.match_type_number',
    'info.outcome.by.runs',
    'info.outcome.by.wickets'
],inplace=True)
```

In [8]:

```python
final_df
```

Out[8]:

	innings	info.dates	info.gender	info.match_type	info.outcome.winner	info.overs	info.player_of_match
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-17]	male	T20	Sri Lanka	20	[DAS Gunaratne]
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-19]	male	T20	Sri Lanka	20	[DAS Gunaratne]
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-22]	male	T20	Australia	20	[A Zampa]
0	[{'1st innings': {'team': 'Hong Kong', 'delive...	[2016-09-05]	male	T20	Hong Kong	20	NaN
0	[{'1st innings': {'team': 'Zimbabwe', 'deliver...	[2016-06-18]	male	T20	Zimbabwe	20	[E Chigumbura]
...	...	...	...	...	...	...	...
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	[2016-03-04]	male	T20	Pakistan	20	[Umar Akmal]
0	[{'1st innings': {'team': 'Bangladesh', 'deliv...	[2016-03-06]	male	T20	India	20	[S Dhawan]
0	[{'1st innings': {'team': 'Netherlands', 'deli...	[2016-02-03]	male	T20	Netherlands	20	[Mudassar Bukhari]
0	[{'1st innings': {'team': 'Australia', 'delive...	[2016-09-06]	male	T20	Australia	20	[GJ Maxwell]
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	[2016-09-09]	male	T20	Australia	20	[GJ Maxwell]

1432 rows × 14 columns

In [9]:

```python
final_df['info.gender'].value_counts()
```

Out[9]:

```
male      966
female    466
Name: info.gender, dtype: int64
```

In [10]:

```python
final_df = final_df[final_df['info.gender'] == 'male']
final_df.drop(columns=['info.gender'],inplace=True)
final_df
```

```
c:\users\92336\appdata\local\programs\python\python39\lib\site-packages\pandas\core\frame.py:4901: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

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

Out[10]:

	innings	info.dates	info.match_type	info.outcome.winner	info.overs	info.player_of_match	info.team
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-17]	T20	Sri Lanka	20	[DAS Gunaratne]	[Australia, Sri Lanka
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-19]	T20	Sri Lanka	20	[DAS Gunaratne]	[Australia, Sri Lanka
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-22]	T20	Australia	20	[A Zampa]	[Australia, Sri Lanka
0	[{'1st innings': {'team': 'Hong Kong', 'delive...	[2016-09-05]	T20	Hong Kong	20	NaN	[Ireland, Hong Kong
0	[{'1st innings': {'team': 'Zimbabwe', 'deliver...	[2016-06-18]	T20	Zimbabwe	20	[E Chigumbura]	[Zimbabwe, India
...	...	...	...	...	...	...	...
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	[2016-03-04]	T20	Pakistan	20	[Umar Akmal]	[Pakistan, Sri Lanka
0	[{'1st innings': {'team': 'Bangladesh', 'deliv...	[2016-03-06]	T20	India	20	[S Dhawan]	[Bangladesh, India
0	[{'1st innings': {'team': 'Netherlands', 'deli...	[2016-02-03]	T20	Netherlands	20	[Mudassar Bukhari]	[United Arab Emirates, Netherlands
0	[{'1st innings': {'team': 'Australia', 'delive...	[2016-09-06]	T20	Australia	20	[GJ Maxwell]	[Sri Lanka, Australia
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	[2016-09-09]	T20	Australia	20	[GJ Maxwell]	[Sri Lanka, Australia

966 rows × 13 columns

In [11]:

```python
final_df['info.match_type'].value_counts()
```

Out[11]:

```
T20    966
Name: info.match_type, dtype: int64
```

In [12]:

```python
final_df['info.overs'].value_counts()
```

Out[12]:

```
20    963
50      3
Name: info.overs, dtype: int64
```

In [13]:

```python
final_df = final_df[final_df['info.overs'] == 20]
final_df.drop(columns=['info.overs','info.match_type'],inplace=True)
final_df
```

```
c:\users\92336\appdata\local\programs\python\python39\lib\site-packages\pandas\core\frame.py:4901: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

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

Out[13]:

	innings	info.dates	info.outcome.winner	info.player_of_match	info.teams	info.toss.decision	info.tos
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-17]	Sri Lanka	[DAS Gunaratne]	[Australia, Sri Lanka]	field	
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-19]	Sri Lanka	[DAS Gunaratne]	[Australia, Sri Lanka]	field	
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-22]	Australia	[A Zampa]	[Australia, Sri Lanka]	field	
0	[{'1st innings': {'team': 'Hong Kong', 'delive...	[2016-09-05]	Hong Kong	NaN	[Ireland, Hong Kong]	bat	Ho
0	[{'1st innings': {'team': 'Zimbabwe', 'deliver...	[2016-06-18]	Zimbabwe	[E Chigumbura]	[Zimbabwe, India]	field	
...	...	...	...	...	...	...	
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	[2016-03-04]	Pakistan	[Umar Akmal]	[Pakistan, Sri Lanka]	field	
0	[{'1st innings': {'team': 'Bangladesh', 'deliv...	[2016-03-06]	India	[S Dhawan]	[Bangladesh, India]	field	
0	[{'1st innings': {'team': 'Netherlands', 'deli...	[2016-02-03]	Netherlands	[Mudassar Bukhari]	[United Arab Emirates, Netherlands]	field	Uni
0	[{'1st innings': {'team': 'Australia', 'delive...	[2016-09-06]	Australia	[GJ Maxwell]	[Sri Lanka, Australia]	field	
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	[2016-09-09]	Australia	[GJ Maxwell]	[Sri Lanka, Australia]	bat	

963 rows × 11 columns

In [14]:

```python
import pickle
pickle.dump(final_df, open('dataset_level1','wb'))
```

In [15]:

```python
matches= pickle.load(open('dataset_level1', 'rb'))
matches
```

Out[15]:

	innings	info.dates	info.outcome.winner	info.player_of_match	info.teams	info.toss.decision	info.tos
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-17]	Sri Lanka	[DAS Gunaratne]	[Australia, Sri Lanka]	field	
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-19]	Sri Lanka	[DAS Gunaratne]	[Australia, Sri Lanka]	field	
0	[{'1st innings': {'team': 'Australia', 'delive...	[2017-02-22]	Australia	[A Zampa]	[Australia, Sri Lanka]	field	
0	[{'1st innings': {'team': 'Hong Kong', 'delive...	[2016-09-05]	Hong Kong	NaN	[Ireland, Hong Kong]	bat	Ho
0	[{'1st innings': {'team': 'Zimbabwe', 'deliver...	[2016-06-18]	Zimbabwe	[E Chigumbura]	[Zimbabwe, India]	field	
...	...	...	...	...	...	...	
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	[2016-03-04]	Pakistan	[Umar Akmal]	[Pakistan, Sri Lanka]	field	
0	[{'1st innings': {'team': 'Bangladesh', 'deliv...	[2016-03-06]	India	[S Dhawan]	[Bangladesh, India]	field	
0	[{'1st innings': {'team': 'Netherlands', 'deli...	[2016-02-03]	Netherlands	[Mudassar Bukhari]	[United Arab Emirates, Netherlands]	field	Uni
0	[{'1st innings': {'team': 'Australia', 'delive...	[2016-09-06]	Australia	[GJ Maxwell]	[Sri Lanka, Australia]	field	
0	[{'1st innings': {'team': 'Sri Lanka', 'delive...	[2016-09-09]	Australia	[GJ Maxwell]	[Sri Lanka, Australia]	bat	

963 rows × 11 columns

In [16]:

```python
matches.iloc[0]['innings'][0]['1st innings']['deliveries']
```

Out[16]:


```
[{0.1: {'batsman': 'AJ Finch',
   'bowler': 'SL Malinga',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {0.2: {'batsman': 'AJ Finch',
   'bowler': 'SL Malinga',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {0.3: {'batsman': 'AJ Finch',
   'bowler': 'SL Malinga',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {0.4: {'batsman': 'M Klinger',
   'bowler': 'SL Malinga',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {0.5: {'batsman': 'M Klinger',
   'bowler': 'SL Malinga',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {0.6: {'batsman': 'M Klinger',
   'bowler': 'SL Malinga',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 3, 'extras': 0, 'total': 3}}},
 {1.1: {'batsman': 'M Klinger',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {1.2: {'batsman': 'M Klinger',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {1.3: {'batsman': 'AJ Finch',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {1.4: {'batsman': 'AJ Finch',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {1.5: {'batsman': 'AJ Finch',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 4, 'extras': 0, 'total': 4}}},
 {1.6: {'batsman': 'AJ Finch',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {2.1: {'batsman': 'M Klinger',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {2.2: {'batsman': 'AJ Finch',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {2.3: {'batsman': 'M Klinger',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {2.4: {'batsman': 'M Klinger',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {2.5: {'batsman': 'M Klinger',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 4, 'extras': 0, 'total': 4}}},
 {2.6: {'batsman': 'M Klinger',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {3.1: {'batsman': 'AJ Finch',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {3.2: {'batsman': 'AJ Finch',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {3.3: {'batsman': 'M Klinger',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {3.4: {'batsman': 'AJ Finch',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {3.5: {'batsman': 'M Klinger',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 4, 'extras': 0, 'total': 4}}},
 {3.6: {'batsman': 'M Klinger',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {4.1: {'batsman': 'M Klinger',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 3, 'extras': 0, 'total': 3}}},
 {4.2: {'batsman': 'AJ Finch',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 3, 'extras': 0, 'total': 3}}},
 {4.3: {'batsman': 'M Klinger',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 4, 'extras': 0, 'total': 4}}},
 {4.4: {'batsman': 'M Klinger',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 4, 'extras': 0, 'total': 4}}},
 {4.5: {'batsman': 'M Klinger',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {4.6: {'batsman': 'AJ Finch',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {5.1: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {5.2: {'batsman': 'M Klinger',
   'bowler': 'S Prasanna',
   'extras': {'wides': 1},
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 1, 'total': 1}}},
 {5.3: {'batsman': 'M Klinger',
   'bowler': 'S Prasanna',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {5.4: {'batsman': 'M Klinger',
   'bowler': 'S Prasanna',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {5.5: {'batsman': 'M Klinger',
   'bowler': 'S Prasanna',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {5.6: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {5.7: {'batsman': 'M Klinger',
   'bowler': 'S Prasanna',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {6.1: {'batsman': 'AJ Finch',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {6.2: {'batsman': 'M Klinger',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {6.3: {'batsman': 'AJ Finch',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {6.4: {'batsman': 'AJ Finch',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {6.5: {'batsman': 'M Klinger',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {6.6: {'batsman': 'AJ Finch',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {7.1: {'batsman': 'M Klinger',
   'bowler': 'S Prasanna',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {7.2: {'batsman': 'M Klinger',
   'bowler': 'S Prasanna',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {7.3: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {7.4: {'batsman': 'M Klinger',
   'bowler': 'S Prasanna',
   'extras': {'byes': 1},
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 1, 'total': 1}}},
 {7.5: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 4, 'extras': 0, 'total': 4}}},
 {7.6: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {8.1: {'batsman': 'AJ Finch',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {8.2: {'batsman': 'M Klinger',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 3, 'extras': 0, 'total': 3}}},
 {8.3: {'batsman': 'AJ Finch',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {8.4: {'batsman': 'M Klinger',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {8.5: {'batsman': 'M Klinger',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {8.6: {'batsman': 'M Klinger',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {9.1: {'batsman': 'M Klinger',
   'bowler': 'S Prasanna',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {9.2: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {9.3: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {9.4: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {9.5: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 6, 'extras': 0, 'total': 6}}},
 {9.6: {'batsman': 'AJ Finch',
   'bowler': 'S Prasanna',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {10.1: {'batsman': 'AJ Finch',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {10.2: {'batsman': 'AJ Finch',
   'bowler': 'PADLR Sandakan',
   'extras': {'legbyes': 1},
   'non_striker': 'M Klinger',
   'runs': {'batsman': 0, 'extras': 1, 'total': 1}}},
 {10.3: {'batsman': 'M Klinger',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0},
   'wicket': {'fielders': ['SL Malinga'],
    'kind': 'caught',
    'player_out': 'M Klinger'}}},
 {10.4: {'batsman': 'TM Head',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {10.5: {'batsman': 'TM Head',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {10.6: {'batsman': 'TM Head',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'AJ Finch',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {11.1: {'batsman': 'AJ Finch',
   'bowler': 'DAS Gunaratne',
   'non_striker': 'TM Head',
   'runs': {'batsman': 6, 'extras': 0, 'total': 6}}},
 {11.2: {'batsman': 'AJ Finch',
   'bowler': 'DAS Gunaratne',
   'non_striker': 'TM Head',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {11.3: {'batsman': 'AJ Finch',
   'bowler': 'DAS Gunaratne',
   'non_striker': 'TM Head',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0},
   'wicket': {'fielders': ['SL Malinga'],
    'kind': 'caught',
    'player_out': 'AJ Finch'}}},
 {11.4: {'batsman': 'TM Head',
   'bowler': 'DAS Gunaratne',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {11.5: {'batsman': 'MC Henriques',
   'bowler': 'DAS Gunaratne',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {11.6: {'batsman': 'TM Head',
   'bowler': 'DAS Gunaratne',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {12.1: {'batsman': 'TM Head',
   'bowler': 'S Prasanna',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {12.2: {'batsman': 'TM Head',
   'bowler': 'S Prasanna',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {12.3: {'batsman': 'MC Henriques',
   'bowler': 'S Prasanna',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {12.4: {'batsman': 'TM Head',
   'bowler': 'S Prasanna',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {12.5: {'batsman': 'TM Head',
   'bowler': 'S Prasanna',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {12.6: {'batsman': 'MC Henriques',
   'bowler': 'S Prasanna',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {13.1: {'batsman': 'MC Henriques',
   'bowler': 'SL Malinga',
   'non_striker': 'TM Head',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {13.2: {'batsman': 'MC Henriques',
   'bowler': 'SL Malinga',
   'extras': {'wides': 1},
   'non_striker': 'TM Head',
   'runs': {'batsman': 0, 'extras': 1, 'total': 1}}},
 {13.3: {'batsman': 'MC Henriques',
   'bowler': 'SL Malinga',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {13.4: {'batsman': 'TM Head',
   'bowler': 'SL Malinga',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {13.5: {'batsman': 'TM Head',
   'bowler': 'SL Malinga',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {13.6: {'batsman': 'MC Henriques',
   'bowler': 'SL Malinga',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {13.7: {'batsman': 'TM Head',
   'bowler': 'SL Malinga',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {14.1: {'batsman': 'TM Head',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {14.2: {'batsman': 'MC Henriques',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'TM Head',
   'runs': {'batsman': 3, 'extras': 0, 'total': 3}}},
 {14.3: {'batsman': 'TM Head',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {14.4: {'batsman': 'MC Henriques',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'TM Head',
   'runs': {'batsman': 6, 'extras': 0, 'total': 6}}},
 {14.5: {'batsman': 'MC Henriques',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'TM Head',
   'runs': {'batsman': 3, 'extras': 0, 'total': 3}}},
 {14.6: {'batsman': 'TM Head',
   'bowler': 'PADLR Sandakan',
   'non_striker': 'MC Henriques',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {15.1: {'batsman': 'MC Henriques',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'TM Head',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0},
   'wicket': {'fielders': ['CK Kapugedera'],
    'kind': 'caught',
    'player_out': 'MC Henriques'}}},
 {15.2: {'batsman': 'TM Head',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {15.3: {'batsman': 'AJ Turner',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'TM Head',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {15.4: {'batsman': 'AJ Turner',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {15.5: {'batsman': 'TM Head',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {15.6: {'batsman': 'TM Head',
   'bowler': 'JRMVB Sanjaya',
   'extras': {'wides': 1},
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 0, 'extras': 1, 'total': 1}}},
 {15.7: {'batsman': 'TM Head',
   'bowler': 'JRMVB Sanjaya',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 6, 'extras': 0, 'total': 6}}},
 {16.1: {'batsman': 'AJ Turner',
   'bowler': 'SL Malinga',
   'non_striker': 'TM Head',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {16.2: {'batsman': 'AJ Turner',
   'bowler': 'SL Malinga',
   'non_striker': 'TM Head',
   'runs': {'batsman': 4, 'extras': 0, 'total': 4}}},
 {16.3: {'batsman': 'AJ Turner',
   'bowler': 'SL Malinga',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {16.4: {'batsman': 'TM Head',
   'bowler': 'SL Malinga',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {16.5: {'batsman': 'AJ Turner',
   'bowler': 'SL Malinga',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {16.6: {'batsman': 'TM Head',
   'bowler': 'SL Malinga',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {17.1: {'batsman': 'TM Head',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {17.2: {'batsman': 'AJ Turner',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'TM Head',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {17.3: {'batsman': 'AJ Turner',
   'bowler': 'KMDN Kulasekara',
   'extras': {'noballs': 1},
   'non_striker': 'TM Head',
   'runs': {'batsman': 0, 'extras': 1, 'total': 1}}},
 {17.4: {'batsman': 'AJ Turner',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {17.5: {'batsman': 'TM Head',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {17.6: {'batsman': 'AJ Turner',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {17.7: {'batsman': 'TM Head',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {18.1: {'batsman': 'AJ Turner',
   'bowler': 'SL Malinga',
   'non_striker': 'TM Head',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {18.2: {'batsman': 'AJ Turner',
   'bowler': 'SL Malinga',
   'non_striker': 'TM Head',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {18.3: {'batsman': 'TM Head',
   'bowler': 'SL Malinga',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {18.4: {'batsman': 'TM Head',
   'bowler': 'SL Malinga',
   'non_striker': 'AJ Turner',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0},
   'wicket': {'fielders': ['S Prasanna'],
    'kind': 'caught',
    'player_out': 'TM Head'}}},
 {18.5: {'batsman': 'AJ Turner',
   'bowler': 'SL Malinga',
   'non_striker': 'JP Faulkner',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0},
   'wicket': {'fielders': ['S Prasanna'],
    'kind': 'caught',
    'player_out': 'AJ Turner'}}},
 {18.6: {'batsman': 'JP Faulkner',
   'bowler': 'SL Malinga',
   'non_striker': 'TD Paine',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1}}},
 {19.1: {'batsman': 'JP Faulkner',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'TD Paine',
   'runs': {'batsman': 4, 'extras': 0, 'total': 4}}},
 {19.2: {'batsman': 'JP Faulkner',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'TD Paine',
   'runs': {'batsman': 0, 'extras': 0, 'total': 0}}},
 {19.3: {'batsman': 'JP Faulkner',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'TD Paine',
   'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
 {19.4: {'batsman': 'JP Faulkner',
   'bowler': 'KMDN Kulasekara',
   'extras': {'wides': 1},
   'non_striker': 'TD Paine',
   'runs': {'batsman': 0, 'extras': 1, 'total': 1}}},
 {19.5: {'batsman': 'JP Faulkner',
   'bowler': 'KMDN Kulasekara',
   'non_striker': 'TD Paine',
   'runs': {'batsman': 1, 'extras': 0, 'total': 1},
   'wicket': {'fielders': ['CK Kapugedera', 'KMDN Kulasekara'],
```

```
  'kind': 'run out',
  'player_out': 'TD Paine'}}},
{19.6: {'batsman': 'JP Faulkner',
  'bowler': 'KMDN Kulasekara',
  'non_striker': 'PJ Cummins',
  'runs': {'batsman': 2, 'extras': 0, 'total': 2}}},
{19.7: {'batsman': 'JP Faulkner',
  'bowler': 'KMDN Kulasekara',
  'non_striker': 'PJ Cummins',
  'runs': {'batsman': 4, 'extras': 0, 'total': 4}}}]
```

In [28]:

```python
count = 1
delivery_df = pd.DataFrame()
for index, row in matches.iterrows():
    if count in [75,108,150,180,268,360,443,458,584,748,982,1052,1111,1226,1345]:
        count+=1
        continue
    count+=1
    ball_of_match = []
    batsman = []
    bowler = []
    runs = []
    player_of_dismissed = []
    teams = []
    batting_team = []
    match_id = []
    city = []
    venue = []
    for ball in row['innings'][0]['1st innings']['deliveries']:
        for key in ball.keys():
            match_id.append(count)
            batting_team.append(row['innings'][0]['1st innings']['team'])
            teams.append(row['info.teams'])
            ball_of_match.append(key)
            batsman.append(ball[key]['batsman'])
            bowler.append(ball[key]['bowler'])
            runs.append(ball[key]['runs']['total'])
            city.append(row['info.city'])
            venue.append(row['info.venue'])
            try:
                player_of_dismissed.append(ball[key]['wicket']['player_out'])
            except:
                player_of_dismissed.append('0')
    loop_df = pd.DataFrame({
            'match_id':match_id,
            'teams':teams,
            'batting_team':batting_team,
            'ball':ball_of_match,
            'batsman':batsman,
            'bowler':bowler,
            'runs':runs,
            'player_dismissed':player_of_dismissed,
            'city':city,
            'venue':venue
        })
    delivery_df = delivery_df.append(loop_df)
```

In [29]:

```python
delivery_df
```

Out[29]:

	match_id	teams	batting_team	ball	batsman	bowler	runs	player_dismissed	city	venu
0	2	[Australia, Sri Lanka]	Australia	0.1	AJ Finch	SL Malinga	0	0	NaN	Melbourn Cricke Groun
1	2	[Australia, Sri Lanka]	Australia	0.2	AJ Finch	SL Malinga	0	0	NaN	Melbourn Cricke Groun
2	2	[Australia, Sri Lanka]	Australia	0.3	AJ Finch	SL Malinga	1	0	NaN	Melbourn Cricke Groun
3	2	[Australia, Sri Lanka]	Australia	0.4	M Klinger	SL Malinga	2	0	NaN	Melbourn Cricke Groun
4	2	[Australia, Sri Lanka]	Australia	0.5	M Klinger	SL Malinga	0	0	NaN	Melbourn Cricke Groun
...	...	...	...	...	...	...	...	...	...	...
121	964	[Sri Lanka, Australia]	Sri Lanka	19.3	SMSM Senanayake	MA Starc	1	0	Colombo	Premadas Stadiun
122	964	[Sri Lanka, Australia]	Sri Lanka	19.4	DM de Silva	MA Starc	0	0	Colombo	Premadas Stadiun
123	964	[Sri Lanka, Australia]	Sri Lanka	19.5	DM de Silva	MA Starc	0	DM de Silva	Colombo	Premadas Stadiun
124	964	[Sri Lanka, Australia]	Sri Lanka	19.6	SMSM Senanayake	MA Starc	2	0	Colombo	Premadas Stadiun
125	964	[Sri Lanka, Australia]	Sri Lanka	19.7	SMSM Senanayake	MA Starc	1	0	Colombo	Premadas Stadiun

115325 rows × 10 columns

In [30]:

```python
def bowl(row):
    for team in row['teams']:
        if team != row['batting_team']:
            return team
```

In [31]:

```python
delivery_df['bowling_team'] = delivery_df.apply(bowl,axis=1)
```

In [32]:

```python
delivery_df.drop(columns='teams',inplace=True)
```

In [33]:

```python
delivery_df['batting_team'].unique()
```

Out[33]:

```
array(['Australia', 'Hong Kong', 'Zimbabwe', 'India', 'Bangladesh',
       'New Zealand', 'South Africa', 'England', 'West Indies', 'Ireland',
       'Afghanistan', 'Pakistan', 'United Arab Emirates', 'Scotland',
       'Oman', 'Papua New Guinea', 'Sri Lanka', 'Netherlands', 'Nepal',
       'Vanuatu', 'Philippines', 'United States of America', 'Germany',
       'Ghana', 'Uganda', 'Kenya', 'Namibia', 'Nigeria', 'Botswana',
       'Guernsey', 'Denmark', 'Jersey', 'Italy', 'Norway', 'Thailand',
       'Malaysia', 'Maldives', 'Singapore', 'Kuwait', 'Bermuda', 'Canada',
       'Cayman Islands', 'Portugal', 'Gibraltar', 'Spain', 'Bhutan',
       'Qatar', 'Iran', 'Belgium', 'Isle of Man', 'Bulgaria', 'Romania'],
      dtype=object)
```

In [34]:

```python
teams = [
    'Australia',
    'India',
    'Bangladesh',
    'Sri Lanka',
    'Pakistan',
    'Afghanistan',
    'New Zealand',
    'South Africa',
    'West Indies',
    'England'
]
```

In [35]:

```python
delivery_df=delivery_df[delivery_df['batting_team'].isin(teams)]
delivery_df=delivery_df[delivery_df['bowling_team'].isin(teams)]
```

In [36]:

```python
delivery_df
```

Out[36]:

	match_id	batting_team	ball	batsman	bowler	runs	player_dismissed	city	venue	bowling_
0	2	Australia	0.1	AJ Finch	SL Malinga	0	0	NaN	Melbourne Cricket Ground	Sri
1	2	Australia	0.2	AJ Finch	SL Malinga	0	0	NaN	Melbourne Cricket Ground	Sri
2	2	Australia	0.3	AJ Finch	SL Malinga	1	0	NaN	Melbourne Cricket Ground	Sri
3	2	Australia	0.4	M Klinger	SL Malinga	2	0	NaN	Melbourne Cricket Ground	Sri
4	2	Australia	0.5	M Klinger	SL Malinga	0	0	NaN	Melbourne Cricket Ground	Sri
...	...	...	...	...	...	...	...	...	...	
121	964	Sri Lanka	19.3	SMSM Senanayake	MA Starc	1	0	Colombo	R Premadasa Stadium	Au
122	964	Sri Lanka	19.4	DM de Silva	MA Starc	0	0	Colombo	R Premadasa Stadium	Au
123	964	Sri Lanka	19.5	DM de Silva	MA Starc	0	DM de Silva	Colombo	R Premadasa Stadium	Au
124	964	Sri Lanka	19.6	SMSM Senanayake	MA Starc	2	0	Colombo	R Premadasa Stadium	Au
125	964	Sri Lanka	19.7	SMSM Senanayake	MA Starc	1	0	Colombo	R Premadasa Stadium	Au

63888 rows × 10 columns

In [37]:

```python
output = delivery_df[['match_id', 'batting_team', 'bowling_team','ball','runs','player
```

In [38]:

```python
output
```

Out[38]:

	match_id	batting_team	bowling_team	ball	runs	player_dismissed	city	venue
0	2	Australia	Sri Lanka	0.1	0	0	NaN	Melbourne Cricket Ground
1	2	Australia	Sri Lanka	0.2	0	0	NaN	Melbourne Cricket Ground
2	2	Australia	Sri Lanka	0.3	1	0	NaN	Melbourne Cricket Ground
3	2	Australia	Sri Lanka	0.4	2	0	NaN	Melbourne Cricket Ground
4	2	Australia	Sri Lanka	0.5	0	0	NaN	Melbourne Cricket Ground
...	...	...	...	...	...	...	...	...
121	964	Sri Lanka	Australia	19.3	1	0	Colombo	R Premadasa Stadium
122	964	Sri Lanka	Australia	19.4	0	0	Colombo	R Premadasa Stadium
123	964	Sri Lanka	Australia	19.5	0	DM de Silva	Colombo	R Premadasa Stadium
124	964	Sri Lanka	Australia	19.6	2	0	Colombo	R Premadasa Stadium
125	964	Sri Lanka	Australia	19.7	1	0	Colombo	R Premadasa Stadium

63888 rows × 8 columns

In [39]:

```python
pickle.dump(output, open('dataset_level2.pkl','wb'))
```

In [ ]: