

✓ Import Libraries


```
import pandas as pd
import json
```

✓ Process Match Results

✓ Reading and transforming data

```
with open('/content/t20_wc_match_results.json') as f:
    data = json.load(f)
```

```
df_match = pd.DataFrame(data[0]['matchSummary'])
df_match.head()
```




	team1	team2	winner	margin	ground	mat
0	Namibia	Sri Lanka	Namibia	55 runs	Geelong	
1	Netherlands	U.A.E.	Netherlands	3 wickets	Geelong	
2	Scotland	West Indies	Scotland	42 runs	Hobart	
3	Ireland	Zimbabwe	Zimbabwe	31 runs	Hobart	
4	Namibia	Netherlands	Netherlands	5 wickets	Geelong	

Next
steps:

[code](#)[df_match](#)[recommended](#)[interactive](#)

```
df_match.shape
```

 (45, 7)

✓ Use scorecard as a match id to link with other tables

```
df_match.rename({'scorecard': 'match_id'}, axis = 1, inplace =
df_match.head())
```

t20_wc_player_info.json X

...

```
1 [
2   {
3     "name": "Michael van Lingen",
4     "team": "Namibia",
5     "battingStyle": "Left hand Bat",
6     "bowlingStyle": "Left arm Medi",
7     "playingRole": "Bowling Allrou",
8     "description": ""
9   },
10  {
11    "name": "Divan la Cock",
12    "team": "Namibia",
13    "battingStyle": "Right hand Ba",
14    "bowlingStyle": "Legbreak",
15    "playingRole": "Opening Batter",
16    "description": ""
17  },
18  {
19    "name": "Jan Nicol Loftie-Eato",
20    "team": "Namibia",
21    "battingStyle": "Left hand Bat",
22    "bowlingStyle": "Right arm Med",
23    "playingRole": "Batter",
24    "description": ""
25  },
26 ]
```



	team1	team2	winner	margin	ground	mat
0	Namibia	Sri Lanka	Namibia	55 runs	Geelong	
1	Netherlands	U.A.E.	Netherlands	3 wickets	Geelong	
2	Scotland	West Indies	Scotland	42 runs	Hobart	
3	Ireland	Zimbabwe	Zimbabwe	31 runs	Hobart	
4	Namibia	Netherlands	Netherlands	5 wickets	Geelong	

Next
steps:

[code](#) [df_match](#)
☒ [recommended](#)
[interactive](#)

- ✓ Create a match ids dictionary that maps team names to a unique match id. This will be useful later on to link with other tables

```
match_ids_dict = {}

for index, row in df_match.iterrows():
    key1 = row['team1'] + ' Vs ' + row['team2']
    key2 = row['team2'] + ' Vs ' + row['team1']
    match_ids_dict[key1] = row['match_id']
    match_ids_dict[key2] = row['match_id']
```

✓ Process Batting Results

✓ Reading and transforming data

```
with open('/content/sample_data/t20-json-files/t20_wc_batting_
data = json.load(f)
all_records = []
for rec in data:
    all_records.extend(rec['battingSummary'])

df_batting = pd.DataFrame(all_records)
df_batting.head(11)
```



	match	teamInnings	battingPos	batsmanName	dismissal
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pra Madush Cham
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shana Pra Madus
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c †Men Karunar
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c Df Sil Pra Madus
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	Gunathi b PWI
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	run (Gunathil †Mer
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	c †Men Theekst
7	Namibia Vs Sri Lanka	Namibia	8	JJ Smit	
8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	c Sr Shik
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendis†	c †Gre W
10	Namibia Vs Sri Lanka	Sri Lanka	3	Dhananjaya de Silva	c Shikon Fry

Next steps:

[code](#) [df_batting](#)

☒ [recommended](#)

[interactive](#)

```
df_batting['out/not_out'] = df_batting.dismissal.apply(lambda x:
df_batting.head(11))
```



	match	teamInnings	battingPos	batsmanName	dismissal
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pra Madush Cham
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shana Pra Madus
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c †Men Karunar
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c Df Sil Pra Madus
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	Gunathi b PWI
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	run (Gunathil †Mer
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	c †Men Theekst
7	Namibia Vs Sri Lanka	Namibia	8	JJ Smit	
8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	c Sr Shik
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendis†	c †Gre W
10	Namibia Vs Sri Lanka	Sri Lanka	3	Dhananjaya de Silva	c Shikon Fry

Next steps:

[code](#) [df_batting](#)

☒ [recommended](#)

[interactive](#)

```
df_batting.drop(columns=["dismissal"], inplace=True)
df_batting.head(10)
```



	match	teamInnings	battingPos	batsmanName	runs	balls
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	3	
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	9	
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	20	
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	26	
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	20	
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	44	
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	0	
7	Namibia Vs Sri Lanka	Namibia	8	JJ Smit	31	
8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	9	
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendis†	6	

Next
steps:

[code](#) [df_batting](#)

☒ [recommended](#)

[interactive](#)

```
df_batting['match_id'] = df_batting['match'].map(match_ids_dic)
df_batting.head()
```



	match	teamInnings	battingPos	batsmanName	runs	balls
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	3	
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	9	
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	20	
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	26	
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	20	

Next
steps:

[code](#) [df_batting](#)☒ recommended[interactive](#)

✓ Cleanup weird characters

```
df_batting['batsmanName'] = df_batting['batsmanName'].apply(lambda x: x.replace(' ', ''))
df_batting['batsmanName'] = df_batting['batsmanName'].apply(lambda x: x.replace(' ', ''))
df_batting.head()
```



	match	teamInnings	battingPos	batsmanName	runs	balls
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	3	
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	9	
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	20	
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	26	
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	20	

Next
steps:

[code](#) [df_batting](#)☒ recommended[interactive](#)

```
df_batting.shape
```



```
(699, 11)
```

✓ Process Bowling Results

✓ Reading and transforming data

```
with open('/content/t20_wc_bowling_summary.json') as f:
    data = json.load(f)
    all_records = []
    for rec in data:
        all_records.extend(rec['bowlingSummary'])
all_records[:2]
```

```
➞ [{"match": 'Namibia Vs Sri Lanka',
    'bowlingTeam': 'Sri Lanka',
    'bowlerName': 'Maheesh Theekshana',
    'overs': '4',
    'maiden': '0',
    'runs': '23',
    'wickets': '1',
    'economy': '5.75',
    '0s': '7',
    '4s': '0',
    '6s': '0',
    'wides': '2',
    'noBalls': '0'},
   {'match': 'Namibia Vs Sri Lanka',
    'bowlingTeam': 'Sri Lanka',
    'bowlerName': 'Dushmantha Chameera',
    'overs': '4',
    'maiden': '0',
    'runs': '39',
    'wickets': '1',
    'economy': '9.75',
    '0s': '6',
    '4s': '3',
    '6s': '1',
    'wides': '2',
    'noBalls': '0'}]
```

```
df_bowling = pd.DataFrame(all_records)
print(df_bowling.shape)
df_bowling.head()
```

 (500, 13)

	match	bowlingTeam	bowlerName	overs	maiden	runs	wickets
0	Namibia Vs Sri Lanka	Sri Lanka	Maheesh Theekshana	4	0	23	1
1	Namibia Vs Sri Lanka	Sri Lanka	Dushmantha Chameera	4	0	39	1
2	Namibia Vs Sri Lanka	Sri Lanka	Pramod Madushan	4	0	37	1
3	Namibia Vs Sri Lanka	Sri Lanka	Chamika Karunaratne	4	0	36	1
4	Namibia Vs Sri Lanka	Sri Lanka	Wanindu Hasaranga de Silva	4	0	27	1

Next steps:

[code](#)[df_bowling](#)[recommended](#)[interactive](#)

```
df_bowling['match_id'] = df_bowling['match'].map(match_ids_dict)
df_bowling.head()
```



	match	bowlingTeam	bowlerName	overs	maiden	runs	wickets
0	Namibia Vs Sri Lanka	Sri Lanka	Maheesh Theekshana	4	0	23	1
1	Namibia Vs Sri Lanka	Sri Lanka	Dushmantha Chameera	4	0	39	1
2	Namibia Vs Sri Lanka	Sri Lanka	Pramod Madushan	4	0	37	1
3	Namibia Vs Sri Lanka	Sri Lanka	Chamika Karunaratne	4	0	36	1
4	Namibia Vs Sri Lanka	Sri Lanka	Wanindu Hasaranga de Silva	4	0	27	1

Next steps:

[code](#)[df_bowling](#)[recommended](#)[interactive](#)

✓ Process Player's Information

✓ Reading and transforming data


```
with open('/content/t20_wc_player_info.json') as f:  
    data = json.load(f)
```

```
df_players = pd.DataFrame(data)
```

```
print(df_players.shape)  
df_players.head(10)
```

↔ (219, 6)

	name	team	battingStyle	bowlingStyle	playing
0	Michael van Lingen	Namibia	Left hand Bat	Left arm Medium	Bo
1	Divan la Cock	Namibia	Right hand Bat	Legbreak	Op
2	Jan Nicol Loftie-Eaton	Namibia	Left hand Bat	Right arm Medium, Legbreak	E
3	Stephan Baard	Namibia	Right hand Bat	Right arm Medium fast	E
4	Gerhard Erasmus(c)	Namibia	Right hand Bat	Right arm Offbreak	Allrou
5	Jan Frylinck	Namibia	Left hand Bat	Left arm Fast medium	Allrou
6	David Wiese	Namibia	Right hand Bat	Right arm Medium fast	Allrou
7	JJ Smit	Namibia	Right hand Bat	Left arm Medium fast	Bo
8	Pathum Nissanka	Sri Lanka	Right hand Bat		Top
9	Kusal Mendis†	Sri Lanka	Right hand Bat	Legbreak	Wicketke

Next
steps:

[code](#) [df_players](#)

[recommended](#)

[interactive](#)

✓ Cleanup weird characters

```
df_players['name'] = df_players['name'].apply(lambda x: x.replace('†', ''))  
df_players['name'] = df_players['name'].apply(lambda x: x.replace('c', ''))  
df_players['name'] = df_players['name'].apply(lambda x: x.replace('†', ''))  
df_players.head(10)
```



	name	team	battingStyle	bowlingStyle	playing
0	Michael van Lingen	Namibia	Left hand Bat	Left arm Medium	Bo Allrou
1	Divan la Cock	Namibia	Right hand Bat	Legbreak	Op E
2	Jan Nicol Loftie-Eaton	Namibia	Left hand Bat	Right arm Medium, Legbreak	E
3	Stephan Baard	Namibia	Right hand Bat	Right arm Medium fast	E
4	Gerhard Erasmus(c)	Namibia	Right hand Bat	Right arm Offbreak	Allrou
5	Jan Frylinck	Namibia	Left hand Bat	Left arm Fast medium	Allrou
6	David Wiese	Namibia	Right hand Bat	Right arm Medium fast	Allrou
7	JJ Smit	Namibia	Right hand Bat	Left arm Medium fast	Bo Allrou

Next
step

8 [code](#) [df_players](#) [Sri Lanka](#) [Right hand Bat](#) [recommended](#) [interactive](#) [Top](#) [E](#)

```
df_players[df_players['team'] == 'India']
```



	name	team	battingStyle	bowlingStyle	play
127	KL Rahul	India	Right hand Bat		