

```
import pandas as pd
import json
import string
import re
```

```
with open("/content/drive/MyDrive/Data Analyst/t20_json.json/t20_wc_match_results.json") as f:
    data = json.load(f)
data
```

```
[{'matchSummary': [{'team1': 'Namibia',
  'team2': 'Sri Lanka',
  'winner': 'Namibia',
  'margin': '55 runs',
  'ground': 'Geelong',
  'matchDate': 'Oct 16, 2022',
  'scorecard': 'T20I # 1823'},
{'team1': 'Netherlands',
  'team2': 'U.A.E.',
  'winner': 'Netherlands',
  'margin': '3 wickets',
  'ground': 'Geelong',
  'matchDate': 'Oct 16, 2022',
  'scorecard': 'T20I # 1825'},
{'team1': 'Scotland',
  'team2': 'West Indies',
  'winner': 'Scotland',
  'margin': '42 runs',
  'ground': 'Hobart',
  'matchDate': 'Oct 17, 2022',
  'scorecard': 'T20I # 1826'},
{'team1': 'Ireland',
  'team2': 'Zimbabwe',
  'winner': 'Zimbabwe',
  'margin': '31 runs',
  'ground': 'Hobart',
  'matchDate': 'Oct 17, 2022',
  'scorecard': 'T20I # 1828'},
{'team1': 'Namibia',
  'team2': 'Netherlands',
  'winner': 'Netherlands',
  'margin': '5 wickets',
  'ground': 'Geelong',
  'matchDate': 'Oct 18, 2022',
  'scorecard': 'T20I # 1830'},
{'team1': 'Sri Lanka',
  'team2': 'U.A.E.',
  'winner': 'Sri Lanka',
  'margin': '79 runs',
  'ground': 'Geelong',
  'matchDate': 'Oct 18, 2022',
  'scorecard': 'T20I # 1832'},
{'team1': 'Ireland',
  'team2': 'Scotland',
  'winner': 'Ireland',
  'margin': '6 wickets',
  'ground': 'Hobart',
  'matchDate': 'Oct 19, 2022',
  'scorecard': 'T20I # 1833'},
{'team1': 'West Indies',
  'team2': 'Zimbabwe',
  'winner': 'West Indies',
  'margin': '31 runs',
  'ground': 'Hobart',
  'matchDate': 'Oct 19, 2022',
  'scorecard': 'T20I # 1834'},
{'team1': 'Netherlands',
  'team2': 'Sri Lanka',
```

```
len(data)
```

```
1
```

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

```
df_match.head()
```

```

    team1    team2    winner    margin    ground    matchDate    scorecard
0   Namibia    Sri Lanka    Namibia    55 runs    Geelong    Oct 16, 2022    T20I # 1823
1  Netherlands      U.A.E.    Netherlands    3 wickets    Geelong    Oct 16, 2022    T20I # 1825
2   Scotland    West Indies    Scotland    42 runs    Hobart    Oct 17, 2022    T20I # 1826
3   Ireland    Zimbabwe    Zimbabwe    31 runs    Hobart    Oct 17, 2022    T20I # 1828
df_match.tail()
```

	team1	team2	winner	margin	ground	matchDate	scorecard
40	Bangladesh	Pakistan	Pakistan	5 wickets	Adelaide	Nov 6, 2022	T20I # 1872
41	India	Zimbabwe	India	71 runs	Melbourne	Nov 6, 2022	T20I # 1873
42	New Zealand	Pakistan	Pakistan	7 wickets	Sydney	Nov 9, 2022	T20I # 1877
43	England	India	England	10 wickets	Adelaide	Nov 10, 2022	T20I # 1878
44	England	Pakistan	England	5 wickets	Melbourne	Nov 13, 2022	T20I # 1879

```
df_match.shape
```

(45, 7)

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

	team1	team2	winner	margin	ground	matchDate	match_id
0	Namibia	Sri Lanka	Namibia	55 runs	Geelong	Oct 16, 2022	T20I # 1823
1	Netherlands	U.A.E.	Netherlands	3 wickets	Geelong	Oct 16, 2022	T20I # 1825
2	Scotland	West Indies	Scotland	42 runs	Hobart	Oct 17, 2022	T20I # 1826
3	Ireland	Zimbabwe	Zimbabwe	31 runs	Hobart	Oct 17, 2022	T20I # 1828
4	Namibia	Netherlands	Netherlands	5 wickets	Geelong	Oct 18, 2022	T20I # 1830

```
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']
```

```
match_ids_dict
```

```
{'Namibia Vs Sri Lanka': 'T20I # 1823',
 'Sri Lanka Vs Namibia': 'T20I # 1823',
 'Netherlands Vs U.A.E.': 'T20I # 1825',
 'U.A.E. Vs Netherlands': 'T20I # 1825',
 'Scotland Vs West Indies': 'T20I # 1826',
 'West Indies Vs Scotland': 'T20I # 1826',
 'Ireland Vs Zimbabwe': 'T20I # 1828',
 'Zimbabwe Vs Ireland': 'T20I # 1828',
 'Namibia Vs Netherlands': 'T20I # 1830',
 'Netherlands Vs Namibia': 'T20I # 1830',
 'Sri Lanka Vs U.A.E.': 'T20I # 1832',
 'U.A.E. Vs Sri Lanka': 'T20I # 1832',
 'Ireland Vs Scotland': 'T20I # 1833',
 'Scotland Vs Ireland': 'T20I # 1833',
 'West Indies Vs Zimbabwe': 'T20I # 1834',
 'Zimbabwe Vs West Indies': 'T20I # 1834',
 'Netherlands Vs Sri Lanka': 'T20I # 1835',
 'Sri Lanka Vs Netherlands': 'T20I # 1835',
 'Namibia Vs U.A.E.': 'T20I # 1836',
 'U.A.E. Vs Namibia': 'T20I # 1836',
 'Ireland Vs West Indies': 'T20I # 1837',
 'West Indies Vs Ireland': 'T20I # 1837',
 'Scotland Vs Zimbabwe': 'T20I # 1838',
 'Zimbabwe Vs Scotland': 'T20I # 1838',
 'Australia Vs New Zealand': 'T20I # 1839',
 'New Zealand Vs Australia': 'T20I # 1839',
```

```
'Afghanistan Vs England': 'T20I # 1840',
'England Vs Afghanistan': 'T20I # 1840',
'Ireland Vs Sri Lanka': 'T20I # 1841',
'Sri Lanka Vs Ireland': 'T20I # 1841',
'India Vs Pakistan': 'T20I # 1842',
'Pakistan Vs India': 'T20I # 1842',
'Bangladesh Vs Netherlands': 'T20I # 1843',
'Netherlands Vs Bangladesh': 'T20I # 1843',
'South Africa Vs Zimbabwe': 'T20I # 1844',
'Zimbabwe Vs South Africa': 'T20I # 1844',
'Australia Vs Sri Lanka': 'T20I # 1845',
'Sri Lanka Vs Australia': 'T20I # 1845',
'England Vs Ireland': 'T20I # 1846',
'Ireland Vs England': 'T20I # 1846',
'Afghanistan Vs New Zealand': 'T20I # 1846a',
'New Zealand Vs Afghanistan': 'T20I # 1846a',
'Bangladesh Vs South Africa': 'T20I # 1847',
'South Africa Vs Bangladesh': 'T20I # 1847',
'India Vs Netherlands': 'T20I # 1848',
'Netherlands Vs India': 'T20I # 1848',
'Pakistan Vs Zimbabwe': 'T20I # 1849',
'Zimbabwe Vs Pakistan': 'T20I # 1849',
'Afghanistan Vs Ireland': 'T20I # 1849a',
'Ireland Vs Afghanistan': 'T20I # 1849a',
'Australia Vs England': 'T20I # 1849b',
'England Vs Australia': 'T20I # 1849b',
'New Zealand Vs Sri Lanka': 'T20I # 1850',
'Sri Lanka Vs New Zealand': 'T20I # 1850',
'Bangladesh Vs Zimbabwe': 'T20I # 1851',
'Zimbabwe Vs Bangladesh': 'T20I # 1851',
'Netherlands Vs Pakistan': 'T20I # 1852',
'Pakistan Vs Netherlands': 'T20I # 1852',
```

### Batting Summery

```
with open("/content/drive/MyDrive/Data Analyst/t20_json.json/t20_wc_batting_summary.json") as f:
    data = json.load(f)

    all_records = []

    for rec in data:
        all_records.extend(rec['battingSummary'])
df_batting = pd.DataFrame(all_records)

df_batting.head(20)
```

	match	teamInnings	battingPos	batsmanName		dismissal	runs	balls	4s	6s	SR
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera		3	6	0	0	50.00
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shanaka b Pramod Madushan		9	9	1	0	100.00
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c †Mendis b Karunaratne		20	12	1	2	166.66
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c DM de Silva b Pramod Madushan		26	24	2	0	108.33
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	c Gunathilaka b PWH de Silva		20	24	0	0	83.33
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	run out (Gunathilaka/†Mendis)		44	28	4	0	157.14
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	c †Mendis b Theekshana		0	1	0	0	0.00

```
df_batting["out/not_out"] = df_batting.dismissal.apply(lambda x:"out" if len(x)>0 else "not_out")
```

```
df_batting.head(11)
```

	match	teamInnings	battingPos	batsmanName		dismissal	runs	balls	4s	6s	SR	out/not_out
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera		3	6	0	0	50.00	out
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shanaka b Pramod Madushan		9	9	1	0	100.00	out
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c †Mendis b Karunaratne		20	12	1	2	166.66	out
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c DM de Silva b Pramod Madushan		26	24	2	0	108.33	out
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	c Gunathilaka b PWH de Silva		20	24	0	0	83.33	out
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	run out (Gunathilaka/†Mendis)		44	28	4	0	157.14	out
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	c †Mendis b Theekshana		0	1	0	0	0.00	out
	Namibia Vs Sri Lanka	Namibia										

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

```
df_batting.head(50)
```

	match	teamInnings	battingPos	batsmanName	dismissal	runs	balls	4s	6s	SR	out/not_out
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera	3	6	0	0	50.00	out
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shanaka b Pramod Madushan	9	9	1	0	100.00	out
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c †Mendis b Karunaratne	20	12	1	2	166.66	out
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c DM de Silva b Pramod Madushan	26	24	2	0	108.33	out
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	c Gunathilaka b PWH de Silva	20	24	0	0	83.33	out
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	run out (Gunathilaka/ †Mendis)	44	28	4	0	157.14	out
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	c †Mendis b Theekshana	0	1	0	0	0.00	out
7	Namibia Vs Sri Lanka	Namibia	8	JJ Smit		31	16	2	2	193.75	not_out
8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	c Smit b Shikongo	9	10	1	0	90.00	out
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendis†	c †Green b Wiese	6	6	0	0	100.00	out
10	Namibia Vs Sri Lanka	Sri Lanka	3	Dhananjaya de Silva	c Shikongo b Frylinck	12	11	1	0	109.09	out
11	Namibia Vs Sri Lanka	Sri Lanka	4	Danushka Gunathilaka	c †Green b Shikongo	0	1	0	0	0.00	out
12	Namibia Vs Sri Lanka	Sri Lanka	5	Bhanuka Rajapaksa	c la Cock b Scholtz	20	21	2	0	95.23	out
13	Namibia Vs Sri Lanka	Sri Lanka	6	Dasun Shanaka(c)	c †Green b Frylinck	29	23	2	1	126.08	out
14	Namibia Vs Sri Lanka	Sri Lanka	7	Wanindu Hasaranga de Silva	c Loftie-Eaton b Scholtz	4	8	0	0	50.00	out
15	Namibia Vs Sri Lanka	Sri Lanka	8	Chamika Karunaratne	c Baard b Smit	5	8	0	0	62.50	out

```
from os import replace
# remove spacial character
df_batting['batsmanName']=df_batting['batsmanName'].apply(lambda x:x.replace('~!@#$$%^&*()†', ''))
df_batting['batsmanName']=df_batting['batsmanName'].apply(lambda x:x.replace('\xa0', ''))
df_batting.head(5)
```

	match	teamInnings	battingPos	batsmanName	dismissal	runs	balls	4s	6s	SR	out/not_out
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera	3	6	0	0	50.00	out
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shanaka b Pramod Madushan	9	9	1	0	100.00	out
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c †Mendis b Karunaratne	20	12	1	2	166.66	out
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c DM de Silva b Pramod Madushan	26	24	2	0	108.33	out

```
df_batting['match']=df_batting['match'].replace('Afghanistan Vs Australia', 'Afghanistan Vs Australia')
df_batting['match']=df_batting['match'].replace('T20I # 1864', 'T20I # 1864')
```

```
df_batting['mach_id']=df_batting['match'].map(match_ids_dict)
df_batting.head(5)
```

	match	teamInnings	battingPos	batsmanName	dismissal	runs	balls	4s	6s	SR	out/not_out	mach_id
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera	3	6	0	0	50.00	out	T20I # 1823
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shanaka b Pramod Madushan	9	9	1	0	100.00	out	T20I # 1823
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c †Mendis b Karunaratne	20	12	1	2	166.66	out	T20I # 1823

```
df_batting.to_csv('fact_bating_summary.csv', index = False)
```