

```
In [1]: import pandas as pd
from code.preprocessing import *
from code.analytics import *
```

```
In [2]: match_results_df = pd.read_json('../data/raw/match_results.json')
match_results_df = cleanse_match_results(match_results_df)
mens_results, womens_results = split_by_gender(match_results_df)
print(mens_results.head(), '\n', womens_results.head())
mens_results.to_csv('../data/clean/mens_results.csv')
womens_results.to_csv('../data/clean/womens_results.csv')
```

	dates	outcome.winner	teams	matchid
0	2017-01-13	Australia	Australia	1000887
1	2017-01-13	Australia	Pakistan	1000887
2	2017-01-15	Pakistan	Australia	1000889
3	2017-01-15	Pakistan	Pakistan	1000889
4	2017-01-19	Australia	Australia	1000891

  

	dates	outcome.winner	teams	matchid
156	2017-01-18	South Africa	Bangladesh	1041523
157	2017-01-18	South Africa	South Africa	1041523
158	2017-01-20	South Africa	Bangladesh	1041525
159	2017-01-20	South Africa	South Africa	1041525
160	2016-11-18	Australia	Australia	1043953

```
In [3]: mens_summary = assemble_yearly_summaries_per_team(mens_results)
mens_summary = mens_summary.loc[mens_summary['Year'] == '2019']
mens_summary.to_csv('../data/results/mens_summary_2019.csv')
print(mens_summary.sort_values(by=['TotalWins'], ascending=False).head())
print(mens_summary.sort_values(by=['Win%'], ascending=False).head())
```

	Year	Team	TotalGames	TotalWins	Win%
0	2019	Australia	23	16	0.695652
0	2019	India	23	15	0.652174
0	2019	England	20	14	0.700000
0	2019	New Zealand	20	13	0.650000
0	2019	West Indies	22	10	0.454545

  

	Year	Team	TotalGames	TotalWins	Win%
0	2019	Netherlands	1	1	1.000000
0	2019	Namibia	4	3	0.750000
0	2019	United States of America	7	5	0.714286
0	2019	England	20	14	0.700000
0	2019	Australia	23	16	0.695652

```
In [4]: womens_summary = assemble_yearly_summaries_per_team(womens_results)
womens_summary = womens_summary.loc[womens_summary['Year'] == '2019']
womens_summary.to_csv('../data/results/womens_summary_2019.csv')
print(womens_summary.sort_values(by=['TotalWins'], ascending=False).head())
print(womens_summary.sort_values(by=['Win%'], ascending=False).head())
```

	Year	Team	TotalGames	TotalWins	Win%
0	2019	Australia	10	10	1.000000
0	2019	India	11	9	0.818182
0	2019	England	11	6	0.545455
0	2019	South Africa	7	3	0.428571
0	2019	Pakistan	6	2	0.333333

  

	Year	Team	TotalGames	TotalWins	Win%
0	2019	Australia	10	10	1.000000
0	2019	India	11	9	0.818182
0	2019	England	11	6	0.545455
0	2019	South Africa	7	3	0.428571
0	2019	Pakistan	6	2	0.333333

```
In [ ]:
```

