

Data_Cleaning

October 23, 2019

```
In [1]: import pandas as pd
```

```
In [2]: matches = pd.read_csv('Desktop/IPL-MSDR-master/ipldata/matches.csv')
        deliveries = pd.read_csv('Desktop/IPL-MSDR-master/ipldata/deliveries.csv')
```

```
In [3]: matches = matches.drop(columns = ['umpire1', 'umpire2', 'umpire3', 'date'])
```

```
In [4]: matches = matches.fillna(value = 'None')
        deliveries = deliveries.fillna(value = 0)
```

```
In [5]: matches.loc[matches.team1 == 'Rising Pune Supergiant', 'team1'] = 'Rising Pune Supergiant'
        matches.loc[matches.team2 == 'Rising Pune Supergiant', 'team2'] = 'Rising Pune Supergiant'
```

```
In [8]: matches.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 756 entries, 0 to 755
Data columns (total 14 columns):
id                756 non-null int64
season            756 non-null int64
city              756 non-null object
team1             756 non-null object
team2             756 non-null object
toss_winner       756 non-null object
toss_decision     756 non-null object
result            756 non-null object
dl_applied        756 non-null int64
winner            756 non-null object
win_by_runs       756 non-null int64
win_by_wickets    756 non-null int64
player_of_match   756 non-null object
venue             756 non-null object
dtypes: int64(5), object(9)
memory usage: 82.8+ KB
```

```
In [9]: deliveries.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 179078 entries, 0 to 179077
Data columns (total 21 columns):
match_id          179078 non-null int64
inning            179078 non-null int64
batting_team      179078 non-null object
bowling_team      179078 non-null object
over              179078 non-null int64
ball              179078 non-null int64
batsman           179078 non-null object
non_striker       179078 non-null object
bowler            179078 non-null object
is_super_over     179078 non-null int64
wide_runs         179078 non-null int64
bye_runs          179078 non-null int64
legbye_runs       179078 non-null int64
noball_runs       179078 non-null int64
penalty_runs      179078 non-null int64
batsman_runs      179078 non-null int64
extra_runs        179078 non-null int64
total_runs        179078 non-null int64
player_dismissed  179078 non-null object
dismissal_kind    179078 non-null object
fielder           179078 non-null object
dtypes: int64(13), object(8)
memory usage: 28.7+ MB
```

```
In [10]: matches.to_csv('Desktop/IPL-MSDR-master/ipldata/new_matches.csv')
```

```
In [11]: deliveries.to_csv('Desktop/IPL-MSDR-master/ipldata/new_deliveries.csv')
```