

Basic Analysis

October 23, 2019

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

```
In [34]: matches = pd.read_csv("../IPL-MSDR/ipldata/original data/matches.csv")
         deliveries = pd.read_csv("../IPL-MSDR/ipldata/original data/deliveries.csv")
```

```
In [6]: matches.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 756 entries, 0 to 755
Data columns (total 18 columns):
id                756 non-null int64
season            756 non-null int64
city              749 non-null object
date              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            752 non-null object
win_by_runs       756 non-null int64
win_by_wickets    756 non-null int64
player_of_match   752 non-null object
venue             756 non-null object
umpire1           754 non-null object
umpire2           754 non-null object
umpire3           119 non-null object
dtypes: int64(5), object(13)
memory usage: 68.0+ KB
```

```
In [7]: 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       8834 non-null object
dismissal_kind          8834 non-null object
fielder                6448 non-null object
dtypes: int64(13), object(8)
memory usage: 23.2+ MB

```

```
In [8]: matches['team1'].unique()
```

```
Out[8]: array(['Sunrisers Hyderabad', 'Mumbai Indians', 'Gujarat Lions',
               'Rising Pune Supergiant', 'Royal Challengers Bangalore',
               'Kolkata Knight Riders', 'Delhi Daredevils', 'Kings XI Punjab',
               'Chennai Super Kings', 'Rajasthan Royals', 'Deccan Chargers',
               'Kochi Tuskers Kerala', 'Pune Warriors', 'Rising Pune Supergiants',
               'Delhi Capitals'], dtype=object)
```

```
In [9]: matches['venue'].unique()
```

```
Out[9]: array(['Rajiv Gandhi International Stadium, Uppal',
               'Maharashtra Cricket Association Stadium',
               'Saurashtra Cricket Association Stadium', 'Holkar Cricket Stadium',
               'M Chinnaswamy Stadium', 'Wankhede Stadium', 'Eden Gardens',
               'Feroz Shah Kotla',
               'Punjab Cricket Association IS Bindra Stadium, Mohali',
               'Green Park', 'Punjab Cricket Association Stadium, Mohali',
               'Sawai Mansingh Stadium', 'MA Chidambaram Stadium, Chepauk',
               'Dr DY Patil Sports Academy', 'Newlands', 'St George's Park',
               'Kingsmead', 'SuperSport Park', 'Buffalo Park',
               'New Wanderers Stadium', 'De Beers Diamond Oval',
               'OUTsurance Oval', 'Brabourne Stadium',
               'Sardar Patel Stadium, Motera', 'Barabati Stadium',
               'Vidarbha Cricket Association Stadium, Jamtha',

```

```
'Himachal Pradesh Cricket Association Stadium', 'Nehru Stadium',
'Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium',
'Subrata Roy Sahara Stadium',
'Shaheed Veer Narayan Singh International Stadium',
'JSCA International Stadium Complex', 'Sheikh Zayed Stadium',
'Sharjah Cricket Stadium', 'Dubai International Cricket Stadium',
'M. A. Chidambaram Stadium', 'Feroz Shah Kotla Ground',
'M. Chinnaswamy Stadium', 'Rajiv Gandhi Intl. Cricket Stadium',
'IS Bindra Stadium', 'ACA-VDCA Stadium'], dtype=object)
```

```
In [11]: matches['city'].unique()
```

```
Out[11]: array(['Hyderabad', 'Pune', 'Rajkot', 'Indore', 'Bangalore', 'Mumbai',
'Kolkata', 'Delhi', 'Chandigarh', 'Kanpur', 'Jaipur', 'Chennai',
'Cape Town', 'Port Elizabeth', 'Durban', 'Centurion',
'East London', 'Johannesburg', 'Kimberley', 'Bloemfontein',
'Ahmedabad', 'Cuttack', 'Nagpur', 'Dharamsala', 'Kochi',
'Visakhapatnam', 'Raipur', 'Ranchi', 'Abu Dhabi', 'Sharjah', nan,
'Mohali', 'Bengaluru'], dtype=object)
```

```
In [28]: len(deliveries['bowler'].unique())
```

```
Out[28]: 405
```

```
In [27]: len(deliveries['batsman'].unique())
```

```
Out[27]: 516
```

```
In [35]: players = set()
for i in range(len(deliveries['match_id'])):
    players.add(deliveries['bowler'][i])
    players.add(deliveries['batsman'][i])
    players.add(deliveries['non_striker'][i])
len(players)
```

```
Out[35]: 559
```