

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

```
In [2]: df = pd.read_csv('premier-league-data.csv')
```

```
In [3]: df.head()
```

Out[3]:

	home_team	away_team	home_goals	away_goals	result	season
0	Sheffield United	Liverpool	1	1	D	2006-2007
1	Arsenal	Aston Villa	1	1	D	2006-2007
2	Everton	Watford	2	1	H	?
3	Newcastle United	Wigan Athletic	2	1	H	2006-2007
4	Portsmouth	Blackburn Rovers	3	0	H	2006-2007

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4560 entries, 0 to 4559
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   home_team   4560 non-null   object
 1   away_team   4560 non-null   object
 2   home_goals  4560 non-null   int64
 3   away_goals  4560 non-null   int64
 4   result      4560 non-null   object
 5   season      4560 non-null   object
dtypes: int64(2), object(4)
memory usage: 213.9+ KB
```

▼ Data Cleaning

▼ Remove invalid values from the season column

```
In [ ]:
```

▼ Identify invalid values in goals scored

```
In [ ]:
```

▼ Replace invalid goals for 0

In []:

▼ **Identify and clean invalid results in the `result` column**

In []:

▼ **Analysis**

▼ **What's the average number of goals per match?**

In []:

▼ **Create a new column `total_goals`**

In []:

▼ **Calculate average goals per season**

In []: `goals_per_season = ...`

▼ **What's the biggest goal difference in a match?**

In []:

▼ **What's the team with most away wins?**

In []:

▼ **What's the team with the most goals scored at home?**

In [7]: `df.groupby('home_team')['home_goals'].sum().sort_values(ascending=False).head()`

```
Out[7]: home_team
Manchester City      497
Manchester United    492
Chelsea              487
Arsenal              465
Liverpool            451
Name: home_goals, dtype: int64
```

▼ **What's the team that received the least amount of goals while playing at home?**

```
In [8]: df.groupby('home_team')['away_goals'].sum().sort_values().head()
```

```
Out[8]: home_team
Charlton Athletic      20
Sheffield United      21
Brighton and Hove Albion  25
Huddersfield Town     25
Cardiff City          31
Name: away_goals, dtype: int64
```

```
In [11]: away_goals_df = df.groupby('home_team')[['result', 'away_goals']].agg(
        {'result': 'size', 'away_goals': 'sum'}
    ).sort_values(
        by=['result', 'away_goals'], ascending=[False, True]
    ).rename(columns={'result': 'games_played', 'away_goals': 'goals_received'})

away_goals_df['ratio'] = away_goals_df['goals_received'] / away_goals_df['game
```

```
In [12]: away_goals_df['ratio']
```

```
Out[12]: home_team
Manchester United      0.644737
Liverpool              0.771930
Manchester City        0.776316
Chelsea                0.785088
Arsenal                0.793860
Tottenham Hotspur     0.938596
Everton                1.035088
West Ham United        1.320574
Stoke City             1.110526
Aston Villa            1.231579
Newcastle United       1.242105
Sunderland             1.242105
West Bromwich Albion   1.374269
Fulham                 1.250000
Swansea City           1.278195
Wigan Athletic         1.398496
Southampton            1.131579
Blackburn Rovers       1.140351
Bolton Wanderers       1.315789
Crystal Palace         1.315789
Hull City              1.473684
Burnley                1.092105
Leicester City         1.105263
Middlesbrough          1.184211
Portsmouth             1.184211
Norwich City           1.250000
Watford                1.328947
Birmingham City       1.000000
Queens Park Rangers    1.228070
Reading                1.315789
AFC Bournemouth        1.491228
Wolverhampton Wanderers 1.631579
Charlton Athletic      1.052632
Sheffield United       1.105263
Brighton and Hove Albion 1.315789
Huddersfield Town      1.315789
Cardiff City           1.631579
Blackpool              1.947368
Derby County           2.263158
Name: ratio, dtype: float64
```

```
In [13]: df.groupby('home_team')[['result', 'away_goals']].agg(  
        {'result': 'size', 'away_goals': 'sum'}  
    ).sort_values(  
        by=['result', 'away_goals'], ascending=[False, True]  
    ).rename(columns={'result': 'games_played', 'away_goals': 'goals_received'})
```

Out[13]:

	games_played	goals_received
home_team		
Manchester United	228	147
Liverpool	228	176
Manchester City	228	177
Chelsea	228	179
Arsenal	228	181
Tottenham Hotspur	228	214
Everton	228	236
West Ham United	209	276
Stoke City	190	211
Aston Villa	190	234
Newcastle United	190	236
Sunderland	190	236
West Bromwich Albion	171	235
Fulham	152	190
Swansea City	133	170
Wigan Athletic	133	186
Southampton	114	129
Blackburn Rovers	114	130
Bolton Wanderers	114	150
Crystal Palace	95	125
Hull City	95	140
Burnley	76	83
Leicester City	76	84
Middlesbrough	76	90
Portsmouth	76	90
Norwich City	76	95
Watford	76	101
Birmingham City	57	57
Queens Park Rangers	57	70
Reading	57	75
AFC Bournemouth	57	85
Wolverhampton Wanderers	57	93
Charlton Athletic	19	20
Sheffield United	19	21

	games_played	goals_received
home_team		
Brighton and Hove Albion	19	25
Huddersfield Town	19	25
Cardiff City	19	31
Blackpool	19	37
Derby County	19	43

- ▼ *What's the team with most goals scored playing as a visitor (away from home)?*

```
In [14]: df.groupby('away_team')['away_goals'].sum().sort_values(ascending=False)
```

```
Out[14]: away_team
Arsenal                377
Manchester United      360
Manchester City         353
Chelsea                351
Liverpool              346
Tottenham Hotspur      335
Everton                251
Aston Villa            212
West Ham United         203
Newcastle United       173
Sunderland             162
West Bromwich Albion   154
Stoke City             150
Swansea City           127
Fulham                 126
Southampton            123
Blackburn Rovers       122
Wigan Athletic         117
Bolton Wanderers       111
Crystal Palace         99
Leicester City         96
Hull City              66
Reading                65
Burnley                64
Watford                62
Portsmouth            61
Norwich City           59
AFC Bournemouth        58
Wolverhampton Wanderers 56
Queens Park Rangers    55
Birmingham City       53
Middlesbrough          48
Blackpool              23
Charlton Athletic      15
Huddersfield Town      12
Cardiff City           12
Brighton and Hove Albion 10
Sheffield United        8
Derby County            8
Name: away_goals, dtype: int64
```

```
In [ ]:
```