

Exploratory Analysis of European League Football Data

We have used two sets of data for our analysis.

The first dataset have been collected from <http://www.datahub.io> (<http://www.datahub.io>)
(<https://datahub.io/collections/football> (<https://datahub.io/collections/football>))

EPL : <https://datahub.io/sports-data/english-premier-league> (<https://datahub.io/sports-data/english-premier-league>)

LaLiga : <https://datahub.io/sports-data/spanish-la-liga> (<https://datahub.io/sports-data/spanish-la-liga>)

Serie A : <https://datahub.io/sports-data/italian-serie-a> (<https://datahub.io/sports-data/italian-serie-a>)

Bundesliga : <https://datahub.io/sports-data/german-bundesliga> (<https://datahub.io/sports-data/german-bundesliga>)

Juventus : <https://datahub.io/sports-data/french-ligue-1> (<https://datahub.io/sports-data/french-ligue-1>)

We have 5 zip files - one for each league

Each zip file unpacks into 9 CSV files (one for each season)

Each row in the CSV files corresponds to one league match

The second dataset is a collection of events during European football league matches.

The dataset has been collected from Kaggle (<https://www.kaggle.com/secareanualin/football-events>)
(<https://www.kaggle.com/secareanualin/football-events>)

Each row of the dataset corresponds to an event. Overall the dataset contains information about close to a million events

The dataset files can be downloaded from :

1. <https://www.kaggle.com/secareanualin/football-events/downloads/events.csv/1>
(<https://www.kaggle.com/secareanualin/football-events/downloads/events.csv/1>)
2. <https://www.kaggle.com/secareanualin/football-events/downloads/ginf.csv/1>
(<https://www.kaggle.com/secareanualin/football-events/downloads/ginf.csv/1>)

A dictionary file explaining the columns in the dataset is available at :

<https://www.kaggle.com/secareanualin/football-events/downloads/dictionary.txt/1>
(<https://www.kaggle.com/secareanualin/football-events/downloads/dictionary.txt/1>)

The objective of this analysis is to :

1. Explore the various features in the data and observe the trend in movement of those features across leagues or across years
2. Identify the features influencing the outcome of a match
3. Build a predictive model and test it

Let's try loading one of the CSV files and examine it's meta-data

```
Out[6]: Index(['Date', 'HomeTeam', 'AwayTeam', 'FTHG', 'FTAG', 'FTR', 'HTHG', 'HTAG',
              'HTR', 'Referee', 'HS', 'AS', 'HST', 'AST', 'HF', 'AF', 'HC', 'AC',
              'HY', 'AY', 'HR', 'AR'],
              dtype='object')
```

We can rename the column headers to be more meaningful, like this :

```
Out[7]: Index(['Date', 'HomeTeam', 'AwayTeam', 'FullTime_HomeTeam_Goals',
              'FullTime_AwayTeam_Goals', 'FullTime_Result', 'HalfTime_HomeTeam_Goals',
              'HalfTime_AwayTeam_Goals', 'HalfTime_Result', 'Referee',
              'HomeTeam_Shots', 'AwayTeam_Shots', 'HomeTeam_ShotsOnTarget',
              'AwayTeam_ShotsOnTarget', 'HomeTeam_FoulsCommitted',
              'AwayTeam_FoulsCommitted', 'HomeTeam_Corners', 'AwayTeam_Corners',
              'HomeTeam_YellowCards', 'AwayTeam_YellowCards', 'HomeTeam_RedCards',
              'AwayTeam_RedCards'],
              dtype='object')
```

We can change the default index and use date as index instead

Lets build a function to load data for all seasons and all leagues in a generic way

We'll start with EPL data first

Loading 2018-2019 data from season-1819_csv.csv

Data Shape : (120, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_ |
|------------|--------------|----------------|-------------------------|-------------------------|-----------|
| Date | | | | | |
| 2018-08-10 | Man United | Leicester | 2 | 1 | |
| 2018-08-11 | Bournemouth | Cardiff | 2 | 0 | |
| 2018-08-11 | Fulham | Crystal Palace | 0 | 2 | |
| 2018-08-11 | Huddersfield | Chelsea | 0 | 3 | |
| 2018-08-11 | Newcastle | Tottenham | 1 | 2 | |

Loading 2017-2018 data from season-1718_csv.csv

Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_ |
|------------|----------------|--------------|-------------------------|-------------------------|-----------|
| Date | | | | | |
| 2017-08-11 | Arsenal | Leicester | 4 | 3 | |
| 2017-08-12 | Brighton | Man City | 0 | 2 | |
| 2017-08-12 | Chelsea | Burnley | 2 | 3 | |
| 2017-08-12 | Crystal Palace | Huddersfield | 0 | 3 | |
| 2017-08-12 | Everton | Stoke | 1 | 0 | |

Loading 2016-2017 data from season-1617_csv.csv

Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|------------|----------------|------------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 2016-08-13 | Burnley | Swansea | 0 | 1 | |
| 2016-08-13 | Crystal Palace | West Brom | 0 | 1 | |
| 2016-08-13 | Everton | Tottenham | 1 | 1 | |
| 2016-08-13 | Hull | Leicester | 2 | 1 | |
| 2016-08-13 | Man City | Sunderland | 2 | 1 | |

Loading 2015-2016 data from season-1516_csv.csv
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|------------|-------------|-------------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 2015-08-08 | Bournemouth | Aston Villa | 0 | 1 | |
| 2015-08-08 | Chelsea | Swansea | 2 | 2 | |
| 2015-08-08 | Everton | Watford | 2 | 2 | |
| 2015-08-08 | Leicester | Sunderland | 4 | 2 | |
| 2015-08-08 | Man United | Tottenham | 1 | 0 | |

Loading 2014-2015 data from season-1415_csv.csv
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|------------|------------|----------------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 2014-08-16 | Arsenal | Crystal Palace | 2 | 1 | |
| 2014-08-16 | Leicester | Everton | 2 | 2 | |
| 2014-08-16 | Man United | Swansea | 1 | 2 | |
| 2014-08-16 | QPR | Hull | 0 | 1 | |
| 2014-08-16 | Stoke | Aston Villa | 0 | 1 | |

Loading 2013-2014 data from season-1314_csv.csv

Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|------------|------------|-------------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 2013-08-17 | Arsenal | Aston Villa | 1 | 3 | |
| 2013-08-17 | Liverpool | Stoke | 1 | 0 | |
| 2013-08-17 | Norwich | Everton | 2 | 2 | |
| 2013-08-17 | Sunderland | Fulham | 0 | 1 | |
| 2013-08-17 | Swansea | Man United | 1 | 4 | |

Loading 2012-2013 data from season-1213_csv.csv

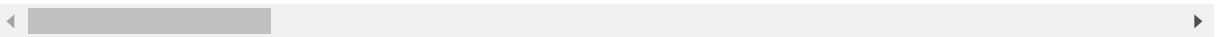
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|------------|-----------|------------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 2012-08-18 | Arsenal | Sunderland | 0 | 0 | |
| 2012-08-18 | Fulham | Norwich | 5 | 0 | |
| 2012-08-18 | Newcastle | Tottenham | 2 | 1 | |
| 2012-08-18 | QPR | Swansea | 0 | 5 | |
| 2012-08-18 | Reading | Stoke | 1 | 1 | |

Loading 2011-2012 data from season-1112_csv.csv

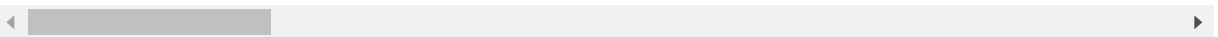
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|------------|-----------|-------------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 2011-08-13 | Blackburn | Wolves | 1 | 2 | |
| 2011-08-13 | Fulham | Aston Villa | 0 | 0 | |
| 2011-08-13 | Liverpool | Sunderland | 1 | 1 | |
| 2011-08-13 | Newcastle | Arsenal | 0 | 0 | |
| 2011-08-13 | QPR | Bolton | 0 | 4 | |



Loading 2010-2011 data from season-1011_csv.csv
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_F |
|------------|-------------|------------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 2010-08-14 | Aston Villa | West Ham | 3 | 0 | |
| 2010-08-14 | Blackburn | Everton | 1 | 0 | |
| 2010-08-14 | Bolton | Fulham | 0 | 0 | |
| 2010-08-14 | Chelsea | West Brom | 6 | 0 | |
| 2010-08-14 | Sunderland | Birmingham | 2 | 2 | |



So data has been loaded correctly. Next, lets concatenate all EPL data together. Except 2018-19 season, which is ongoing, all other datasets have 380 rows. So final dataset will have $8 \times 380 + 120$ i.e. 3160 rows

EPL data loaded fully - Shape: (3160, 21)

Lets repeat the same process for other leagues

Let's try loading La Liga data next. First we'll try to load and examine one of the files

Loading 2018-2019 data from season-1819_csv.csv

Data Shape : (100, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 17/08/18 | Betis | Levante | 0 | 3 | |
| 17/08/18 | Girona | Valladolid | 0 | 0 | |
| 18/08/18 | Barcelona | Alaves | 3 | 0 | |
| 18/08/18 | Celta | Espanol | 1 | 1 | |
| 18/08/18 | Villarreal | Sociedad | 1 | 2 | |



So data has similar layout as EPL data. Lets load data for all the seasons

Loading 2018-2019 data from season-1819_csv.csv

Data Shape : (100, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 17/08/18 | Betis | Levante | 0 | 3 | |
| 17/08/18 | Girona | Valladolid | 0 | 0 | |
| 18/08/18 | Barcelona | Alaves | 3 | 0 | |
| 18/08/18 | Celta | Espanol | 1 | 1 | |
| 18/08/18 | Villarreal | Sociedad | 1 | 2 | |

Loading 2017-2018 data from season-1718_csv.csv

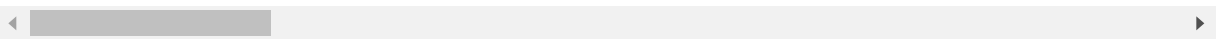
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 18/08/17 | Leganes | Alaves | 1 | 0 | |
| 18/08/17 | Valencia | Las Palmas | 1 | 0 | |
| 19/08/17 | Celta | Sociedad | 2 | 3 | |
| 19/08/17 | Girona | Ath Madrid | 2 | 2 | |
| 19/08/17 | Sevilla | Espanol | 1 | 1 | |

Loading 2016-2017 data from season-1617_csv.csv

Data Shape : (380, 21)

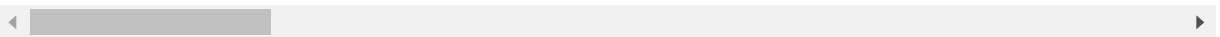
| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-----------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 19/08/16 | La Coruna | Eibar | 2 | 1 | |
| 19/08/16 | Malaga | Osasuna | 1 | 1 | |
| 20/08/16 | Barcelona | Betis | 6 | 2 | |
| 20/08/16 | Granada | Villarreal | 1 | 1 | |
| 20/08/16 | Sevilla | Espanol | 6 | 4 | |



Loading 2015-2016 data from season-1516_csv.csv

Data Shape : (380, 21)

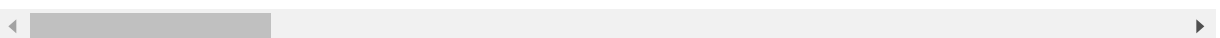
| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 21/08/15 | Malaga | Sevilla | 0 | 0 | |
| 22/08/15 | Ath Madrid | Las Palmas | 1 | 0 | |
| 22/08/15 | Espanol | Getafe | 1 | 0 | |
| 22/08/15 | La Coruna | Sociedad | 0 | 0 | |
| 22/08/15 | Vallecano | Valencia | 0 | 0 | |



Loading 2014-2015 data from season-1415_csv.csv

Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-----------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 23/08/14 | Almeria | Espanol | 1 | 1 | |
| 23/08/14 | Granada | La Coruna | 2 | 1 | |
| 23/08/14 | Malaga | Ath Bilbao | 1 | 0 | |
| 23/08/14 | Sevilla | Valencia | 1 | 1 | |
| 24/08/14 | Barcelona | Elche | 3 | 0 | |



Loading 2013-2014 data from season-1314_csv.csv
 Data Shape : (380, 21)

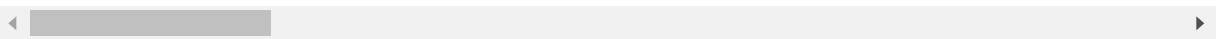
| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 17/08/13 | Sociedad | Getafe | 2 | 0 | |
| 17/08/13 | Valencia | Malaga | 1 | 0 | |
| 17/08/13 | Valladolid | Ath Bilbao | 1 | 2 | |
| 18/08/13 | Barcelona | Levante | 7 | 0 | |
| 18/08/13 | Osasuna | Granada | 1 | 2 | |

Loading 2012-2013 data from season-1213_csv.csv
 Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 18/08/12 | Celta | Malaga | 0 | 1 | |
| 18/08/12 | Mallorca | Espanol | 2 | 1 | |
| 18/08/12 | Sevilla | Getafe | 2 | 1 | |
| 19/08/12 | Ath Bilbao | Betis | 3 | 5 | |
| 19/08/12 | Barcelona | Sociedad | 5 | 1 | |

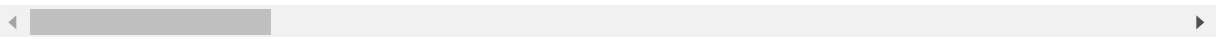
Loading 2011-2012 data from season-1112_csv.csv
 Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|-----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 27/08/11 | Granada | Betis | 0 | 1 | |
| 27/08/11 | Sp Gijon | Sociedad | 1 | 2 | |
| 27/08/11 | Valencia | Santander | 4 | 3 | |
| 28/08/11 | Ath Bilbao | Vallecano | 1 | 1 | |
| 28/08/11 | Ath Madrid | Osasuna | 0 | 0 | |



Loading 2010-2011 data from season-1011_csv.csv
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-----------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 28/08/10 | Hercules | Ath Bilbao | 0 | 1 | |
| 28/08/10 | Levante | Sevilla | 1 | 4 | |
| 28/08/10 | Malaga | Valencia | 1 | 3 | |
| 29/08/10 | Espanol | Getafe | 3 | 1 | |
| 29/08/10 | La Coruna | Zaragoza | 0 | 0 | |



La Liga data fully loaded - Shape: (3140, 21)

Next we'll try to load Serie A data. To begin with, we'll try to load one of the files and see if the data layout meets our expectations

Loading 2018-2019 data from season-1819_csv.csv

Data Shape : (100, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 18/08/18 | Chievo | Juventus | 2 | 3 | |
| 18/08/18 | Lazio | Napoli | 1 | 2 | |
| 19/08/18 | Bologna | Spal | 0 | 1 | |
| 19/08/18 | Empoli | Cagliari | 2 | 0 | |
| 19/08/18 | Parma | Udinese | 2 | 2 | |



So data has similar layout as EPL or La Liga. Lets load data for all the 9 seasons

Loading 2018-2019 data from season-1819_csv.csv

Data Shape : (100, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 18/08/18 | Chievo | Juventus | 2 | 3 | |
| 18/08/18 | Lazio | Napoli | 1 | 2 | |
| 19/08/18 | Bologna | Spal | 0 | 1 | |
| 19/08/18 | Empoli | Cagliari | 2 | 0 | |
| 19/08/18 | Parma | Udinese | 2 | 2 | |

Loading 2017-2018 data from season-1718_csv.csv

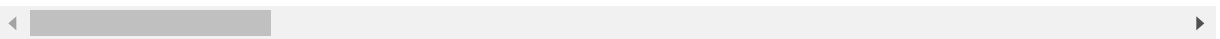
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 19/08/17 | Juventus | Cagliari | 3 | 0 | |
| 19/08/17 | Verona | Napoli | 1 | 3 | |
| 20/08/17 | Atalanta | Roma | 0 | 1 | |
| 20/08/17 | Bologna | Torino | 1 | 1 | |
| 20/08/17 | Crotone | Milan | 0 | 3 | |

Loading 2016-2017 data from season-1617_csv.csv

Data Shape : (380, 21)

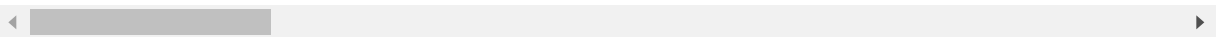
| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 20/08/16 | Juventus | Fiorentina | 2 | 1 | |
| 20/08/16 | Roma | Udinese | 4 | 0 | |
| 21/08/16 | Atalanta | Lazio | 3 | 4 | |
| 21/08/16 | Bologna | Crotone | 1 | 0 | |
| 21/08/16 | Chievo | Inter | 2 | 0 | |



Loading 2015-2016 data from season-1516_csv.csv

Data Shape : (380, 21)

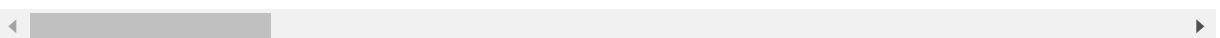
| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 22/08/15 | Lazio | Bologna | 2 | 1 | |
| 22/08/15 | Verona | Roma | 1 | 1 | |
| 23/08/15 | Empoli | Chievo | 1 | 3 | |
| 23/08/15 | Fiorentina | Milan | 2 | 0 | |
| 23/08/15 | Frosinone | Torino | 1 | 2 | |



Loading 2014-2015 data from season-1415_csv.csv

Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 30/08/14 | Chievo | Juventus | 0 | 1 | |
| 30/08/14 | Roma | Fiorentina | 2 | 0 | |
| 31/08/14 | Atalanta | Verona | 0 | 0 | |
| 31/08/14 | Cesena | Parma | 1 | 0 | |
| 31/08/14 | Genoa | Napoli | 1 | 2 | |



Loading 2013-2014 data from season-1314_csv.csv

Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-----------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 24/08/13 | Sampdoria | Juventus | 0 | 1 | |
| 24/08/13 | Verona | Milan | 2 | 1 | |
| 25/08/13 | Cagliari | Atalanta | 2 | 1 | |
| 25/08/13 | Inter | Genoa | 2 | 0 | |
| 25/08/13 | Lazio | Udinese | 2 | 1 | |

Loading 2012-2013 data from season-1213_csv.csv

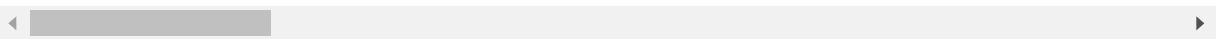
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 25/08/12 | Fiorentina | Udinese | 2 | 1 | |
| 25/08/12 | Juventus | Parma | 2 | 0 | |
| 26/08/12 | Atalanta | Lazio | 0 | 1 | |
| 26/08/12 | Chievo | Bologna | 2 | 0 | |
| 26/08/12 | Genoa | Cagliari | 2 | 0 | |

Loading 2011-2012 data from season-1112_csv.csv

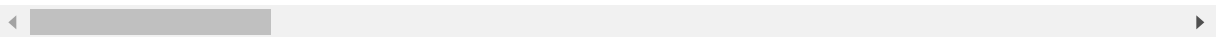
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 09/09/11 | Milan | Lazio | 2 | 2 | |
| 10/09/11 | Cesena | Napoli | 1 | 3 | |
| 11/09/11 | Catania | Siena | 0 | 0 | |
| 11/09/11 | Chievo | Novara | 2 | 2 | |
| 11/09/11 | Fiorentina | Bologna | 2 | 0 | |



Loading 2010-2011 data from season-1011_csv.csv
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|------------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 28/08/10 | Roma | Cesena | 0 | 0 | |
| 28/08/10 | Udinese | Genoa | 0 | 1 | |
| 29/08/10 | Bari | Juventus | 1 | 0 | |
| 29/08/10 | Chievo | Catania | 2 | 1 | |
| 29/08/10 | Fiorentina | Napoli | 1 | 1 | |



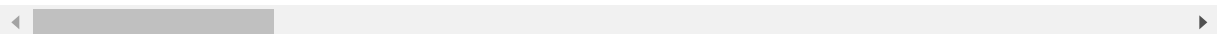
Serie A data loaded fully - shape : (3140, 21)

Next we'll try to load Bundesliga data. To begin with, we'll try to load one of the files and see if the data layout meets our expectations

Loading 2018-2019 data from season-1819_csv.csv

Data Shape : (81, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|--------------------|---------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 24/08/18 | Bayern Munich | Hoffenheim | 3 | 1 | |
| 25/08/18 | Fortuna Dusseldorf | Augsburg | 1 | 2 | |
| 25/08/18 | Freiburg | Ein Frankfurt | 0 | 2 | |
| 25/08/18 | Hertha | Nurnberg | 1 | 0 | |
| 25/08/18 | M'gladbach | Leverkusen | 2 | 0 | |



So data has the same layout as we expected. Lets load data for all the 9 seasons

Loading 2018-2019 data from season-1819_csv.csv

Data Shape : (81, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|--------------------|---------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 24/08/18 | Bayern Munich | Hoffenheim | 3 | 1 | |
| 25/08/18 | Fortuna Dusseldorf | Augsburg | 1 | 2 | |
| 25/08/18 | Freiburg | Ein Frankfurt | 0 | 2 | |
| 25/08/18 | Hertha | Nurnberg | 1 | 0 | |
| 25/08/18 | M'gladbach | Leverkusen | 2 | 0 | |

Loading 2017-2018 data from season-1718_csv.csv

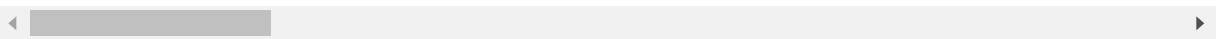
Data Shape : (306, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|---------------|---------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 18/08/17 | Bayern Munich | Leverkusen | 3 | 1 | |
| 19/08/17 | Hamburg | Augsburg | 1 | 0 | |
| 19/08/17 | Hertha | Stuttgart | 2 | 0 | |
| 19/08/17 | Hoffenheim | Werder Bremen | 1 | 0 | |
| 19/08/17 | Mainz | Hannover | 0 | 1 | |

Loading 2016-2017 data from season-1617_csv.csv

Data Shape : (306, 21)

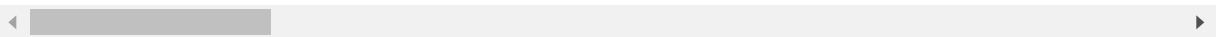
| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|---------------|---------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 26/08/16 | Bayern Munich | Werder Bremen | 6 | 0 | |
| 27/08/16 | Augsburg | Wolfsburg | 0 | 2 | |
| 27/08/16 | Dortmund | Mainz | 2 | 1 | |
| 27/08/16 | Ein Frankfurt | Schalke 04 | 1 | 0 | |
| 27/08/16 | FC Koln | Darmstadt | 2 | 0 | |



Loading 2015-2016 data from season-1516_csv.csv

Data Shape : (306, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|---------------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 14/08/15 | Bayern Munich | Hamburg | 5 | 0 | |
| 15/08/15 | Augsburg | Hertha | 0 | 1 | |
| 15/08/15 | Darmstadt | Hannover | 2 | 2 | |
| 15/08/15 | Dortmund | M'gladbach | 4 | 0 | |
| 15/08/15 | Leverkusen | Hoffenheim | 2 | 1 | |



Loading 2014-2015 data from season-1415_csv.csv

Data Shape : (306, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|---------------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 22/08/14 | Bayern Munich | Wolfsburg | 2 | 1 | |
| 23/08/14 | Dortmund | Leverkusen | 0 | 2 | |
| 23/08/14 | Ein Frankfurt | Freiburg | 1 | 0 | |
| 23/08/14 | FC Koln | Hamburg | 0 | 0 | |
| 23/08/14 | Hannover | Schalke 04 | 2 | 1 | |



Loading 2013-2014 data from season-1314_csv.csv

Data Shape : (306, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTir |
|----------|---------------|---------------|-------------------------|-------------------------|---------|
| Date | | | | | |
| 09/08/13 | Bayern Munich | M'gladbach | 3 | 1 | |
| 10/08/13 | Augsburg | Dortmund | 0 | 4 | |
| 10/08/13 | Braunschweig | Werder Bremen | 0 | 1 | |
| 10/08/13 | Hannover | Wolfsburg | 2 | 0 | |
| 10/08/13 | Hertha | Ein Frankfurt | 6 | 1 | |

Loading 2012-2013 data from season-1213_csv.csv

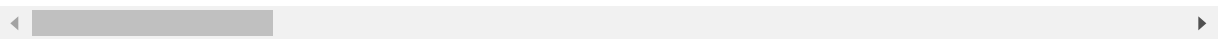
Data Shape : (306, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------------|--------------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 24/08/12 | Dortmund | Werder Bremen | 2 | 1 | |
| 25/08/12 | Augsburg | Fortuna Dusseldorf | 0 | 2 | |
| 25/08/12 | Ein Frankfurt | Leverkusen | 2 | 1 | |
| 25/08/12 | Freiburg | Mainz | 1 | 1 | |
| 25/08/12 | Greuther Furth | Bayern Munich | 0 | 3 | |

Loading 2011-2012 data from season-1112_csv.csv

Data Shape : (306, 21)

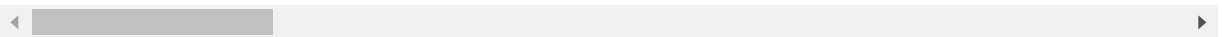
| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 05/08/11 | Dortmund | Hamburg | 3 | 1 | |
| 06/08/11 | Augsburg | Freiburg | 2 | 2 | |
| 06/08/11 | FC Koln | Wolfsburg | 0 | 3 | |
| 06/08/11 | Hannover | Hoffenheim | 2 | 1 | |
| 06/08/11 | Hertha | Nurnberg | 0 | 1 | |



Loading 2010-2011 data from season-1011_csv.csv

Data Shape : (306, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTi |
|----------|---------------|----------------|-------------------------|-------------------------|--------|
| Date | | | | | |
| 20/08/10 | Bayern Munich | Wolfsburg | 2 | 1 | |
| 21/08/10 | FC Koln | Kaiserslautern | 1 | 3 | |
| 21/08/10 | Freiburg | St Pauli | 1 | 3 | |
| 21/08/10 | Hamburg | Schalke 04 | 2 | 1 | |
| 21/08/10 | Hannover | Ein Frankfurt | 2 | 1 | |



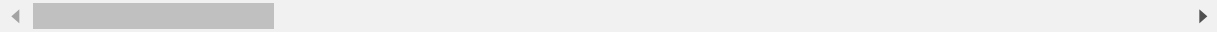
Bundesliga data fully loaded - Shape : (2529, 21)

Next we'll try to load Ligue One data. To begin with, we'll try to load one of the files and see if the data layout meets our expectations

Loading 2018-2019 data from season-1819_csv.csv

Data Shape : (110, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-------------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 10/08/18 | Marseille | Toulouse | 4 | 0 | |
| 11/08/18 | Angers | Nimes | 3 | 4 | |
| 11/08/18 | Lille | Rennes | 3 | 1 | |
| 11/08/18 | Montpellier | Dijon | 1 | 2 | |
| 11/08/18 | Nantes | Monaco | 1 | 3 | |



So data shape is consistent with other league data. Let's proceed with loading data for all the 9 seasons

Loading 2018-2019 data from season-1819_csv.csv

Data Shape : (110, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-------------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 10/08/18 | Marseille | Toulouse | 4 | 0 | |
| 11/08/18 | Angers | Nimes | 3 | 4 | |
| 11/08/18 | Lille | Rennes | 3 | 1 | |
| 11/08/18 | Montpellier | Dijon | 1 | 2 | |
| 11/08/18 | Nantes | Monaco | 1 | 3 | |

Loading 2017-2018 data from season-1718_csv.csv

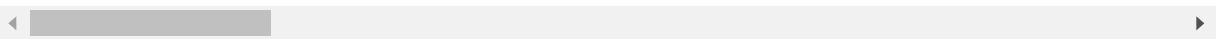
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-------------|------------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 04/08/17 | Monaco | Toulouse | 3 | 2 | |
| 05/08/17 | Lyon | Strasbourg | 4 | 0 | |
| 05/08/17 | Metz | Guingamp | 1 | 3 | |
| 05/08/17 | Montpellier | Caen | 1 | 0 | |
| 05/08/17 | Paris SG | Amiens | 2 | 0 | |

Loading 2016-2017 data from season-1617_csv.csv

Data Shape : (380, 21)

| Date | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|------------|-------------------------|-------------------------|----------|
| 12/08/16 | Bastia | Paris SG | 0 | 1 | |
| 12/08/16 | Monaco | Guingamp | 2 | 2 | |
| 13/08/16 | Bordeaux | St Etienne | 3 | 2 | |
| 13/08/16 | Caen | Lorient | 3 | 2 | |
| 13/08/16 | Dijon | Nantes | 0 | 1 | |



Loading 2015-2016 data from season-1516_csv.csv

Data Shape : (380, 21)

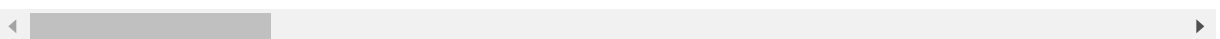
| Date | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-------------|----------|-------------------------|-------------------------|----------|
| 07/08/15 | Lille | Paris SG | 0 | 1 | |
| 08/08/15 | Bastia | Rennes | 2 | 1 | |
| 08/08/15 | Marseille | Caen | 0 | 1 | |
| 08/08/15 | Montpellier | Angers | 0 | 2 | |
| 08/08/15 | Nantes | Guingamp | 1 | 0 | |



Loading 2014-2015 data from season-1415_csv.csv

Data Shape : (380, 21)

| Date | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-----------------------------|------------|-------------------------|-------------------------|----------|
| 08/08/14 | Reims | Paris SG | 2 | 2 | |
| 09/08/14 | Bastia | Marseille | 3 | 3 | |
| 09/08/14 | Evian Thonon Gaillard | Caen | 0 | 3 | |
| 09/08/14 | Guingamp | St Etienne | 0 | 2 | |
| 09/08/14 | Lille | Metz | 0 | 0 | |



Loading 2013-2014 data from season-1314_csv.csv

Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-----------------------------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 09/08/13 | Montpellier | Paris SG | 1 | 1 | |
| 10/08/13 | Bordeaux | Monaco | 0 | 2 | |
| 10/08/13 | Evian Thonon Gaillard | Sochaux | 1 | 1 | |
| 10/08/13 | Lille | Lorient | 1 | 0 | |
| 10/08/13 | Lyon | Nice | 4 | 0 | |

Loading 2012-2013 data from season-1213_csv.csv

Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|-----------------------------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 10/08/12 | Montpellier | Toulouse | 1 | 1 | |
| 11/08/12 | Evian Thonon Gaillard | Bordeaux | 2 | 3 | |
| 11/08/12 | Nancy | Brest | 1 | 0 | |
| 11/08/12 | Nice | Ajaccio | 0 | 1 | |
| 11/08/12 | Paris SG | Lorient | 2 | 2 | |

Loading 2011-2012 data from season-1112_csv.csv

Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|----------|-------------|-----------------------------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 06/08/11 | Ajaccio | Toulouse | 0 | 2 | |
| 06/08/11 | Brest | Evian Thonon Gaillard | 2 | 2 | |
| 06/08/11 | Caen | Valenciennes | 1 | 0 | |
| 06/08/11 | Marseille | Sochaux | 2 | 2 | |
| 06/08/11 | Montpellier | Auxerre | 3 | 1 | |

Loading 2010-2011 data from season-1011_csv.csv
Data Shape : (380, 21)

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|----------|-----------|--------------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 07/08/10 | Auxerre | Lorient | 2 | 2 | |
| 07/08/10 | Lens | Nancy | 1 | 2 | |
| 07/08/10 | Lyon | Monaco | 0 | 0 | |
| 07/08/10 | Marseille | Caen | 1 | 2 | |
| 07/08/10 | Nice | Valenciennes | 0 | 0 | |

Ligue1 data loaded completely - Shape: (3150, 21)

Lets append all the league data into a single dataframe

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|------------|-------------|----------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 2010-08-14 | Aston Villa | West Ham | 3 | 0 | |
| 2010-08-14 | Blackburn | Everton | 1 | 0 | |

Sanitizing/Cleansing the data

Lets check if the data has duplicate rows, if so, remove duplicates from the data

Good news !!! There are no duplicate rows in the data

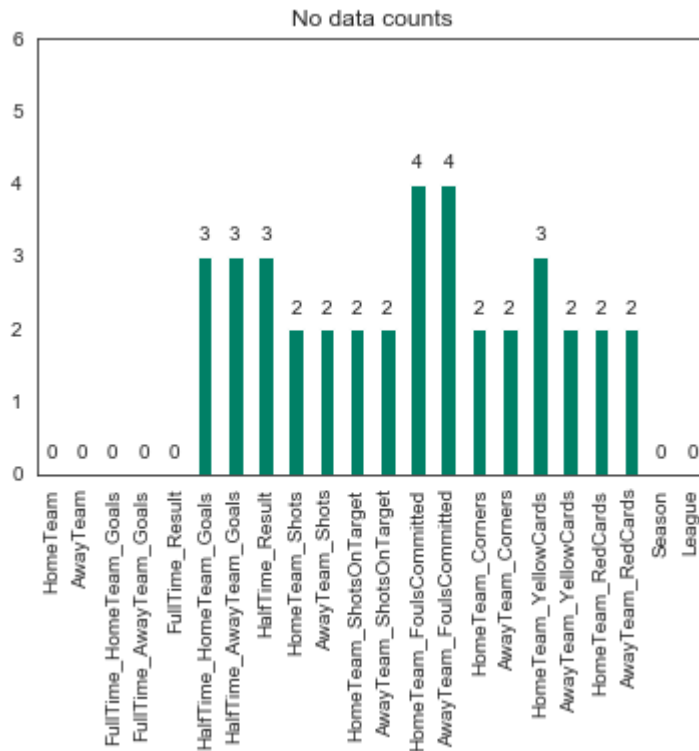
So the data has no duplicate rows. Next we need to check if some of the columns have no values/junk values

```

Full Time Home-Team Goals (Unique Values) : : [ 3  1  0  6  2  4  5  7  8  9
10]
Full Time Away-Team Goals (Unique Values) : : [0  2  4  1  3  6  5  7  8  9]
Full Time Result (Unique Values) : : ['H' 'D' 'A']
Half Time Home-Team Goals (Unique Values) : : [ 2.  1.  0.  3.  4.  5.  6. n
an]
Half Time Away-Team Goals (Unique Values) : : [ 0.  3.  2.  1.  4.  5. nan]
Half Time Result (Unique Values) : : ['H' 'D' 'A' nan]
Home Team Shots (Unique Values) : : [23.  7. 13. 18.  6. 22. 11. 26. 10. 14.
15. 17. 16. 20.  8. 12.  5.  9.
27. 24. 19. 21.  1. 25. 32.  2.  3.  4. 28. 35. 30. 33. 31. 39. 29. 43.
34. 37. 36. nan 38.]
Away Team Shots (Unique Values) : : [12. 17. 10. 13. 11.  9. 14.  7.  3. 15.
 4. 16.  5.  6.  8.  2. 20. 19.
24. 18. 21. 23. 25. 22. 26.  0. 30.  1. 27. 29. 28. 39. 35. nan 31.]
Home Team Shots on Target (Unique Values) : : [11.  2.  9. 13. 18.  6.  7.
 4. 10. 16.  5. 12.  8. 14.  3. 15. 17.  1.
 0. 19. 21. 20. 24. nan]
Away Team Shots on Target (Unique Values) : : [ 2. 12.  7.  4.  6.  3.  1.
 8.  9.  5. 10. 11.  0. 14. 13. 20. 16. 18.
17. 19. 15. nan]
Home Team Fouls Committed (Unique Values) : : [15. 19. 12. 10. 13.  8. 17.
 9. 18. 16. 11.  7.  5.  6. 20. 14.  3. 21.
 4. 23.  2. 22. 24. 25. 31. 30. 29. 26. 27. 28. 33.  1. nan 32.]
Away Team Fouls Committed (Unique Values) : : [15. 14. 13. 10. 16. 11.  5.
 3.  4. 12. 18.  7.  6. 17.  8.  9. 21. 19.
 2. 20. 24.  1. 22. 23. 26. 27. 28. 25. 30. 29.  0. nan 32. 31.]
Home Team Corners (Unique Values) : : [16.  1.  4.  3. 10.  6.  5.  9.  8.
 0.  7. 14. 11.  2. 15. 12. 17. 13.
19. 18. 20. 21. nan 22.]
Away Team Corners (Unique Values) : : [ 7.  3.  8.  1.  6.  4.  5. 11.  2. 1
 2.  0. 16. 10.  9. 14. 13. 15. 17.
19. nan 18.]
Home Team Yellow Cards (Unique Values) : : [ 1.  2.  3.  0.  7.  4.  5.  6.
 8. nan]
Away Team Yellow Cards (Unique Values) : : [ 2.  1.  3.  0.  4.  6.  5.  7.
 8.  9. nan]
Home Team Red Cards (Unique Values) : : [ 0.  1.  2.  3. nan]
Away Team Red Cards (Unique Values) : : [ 0.  1.  2.  3. nan]

```

Clearly there are rows with missing data. Let's get a count of rows with missing information for each column



So we have a few rows with null values for certain columns

Lets try populating the null values now

We can assume that Half Time Results/Goals were in line with Full Time results/goals whenever Half Time data is empty So $\text{Half Time Goals} = \text{Full Time Goals} / 2$

When data is missing we can assume that total no of shots/shots on target is same as the number of goals

Fouls committed, corners and red/yellow cards missing data can be populated based on average number of fouls committed/red or yellow cards in other matches involving these teams

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|-----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 23/09/12 | Cagliari | Roma | 0 | 3 | |
| 18/09/11 | Lyon | Marseille | 2 | 0 | |
| 17/12/11 | Caen | Nancy | 1 | 2 | |
| 16/04/17 | Bastia | Lyon | 0 | 3 | |

So, foul information is missing for the following matches :

1. Lyon vs Marseille
2. Caen vs Nancy
3. Bastia vs Lyon
4. Cagliari vs Roma

We can use other matches involving these teams to populate the missing data

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 23/09/12 | Cagliari | Roma | 0 | 3 | |
| 02/03/15 | Roma | Juventus | 1 | 1 | |
| 16/04/17 | Bastia | Lyon | 0 | 3 | |

Card information is missing for the following matches :

1. Cagliari vs Roma
2. Roma vs Juventus
3. Bastia vs Lyon

We can use other matches involving these teams to populate the missing data

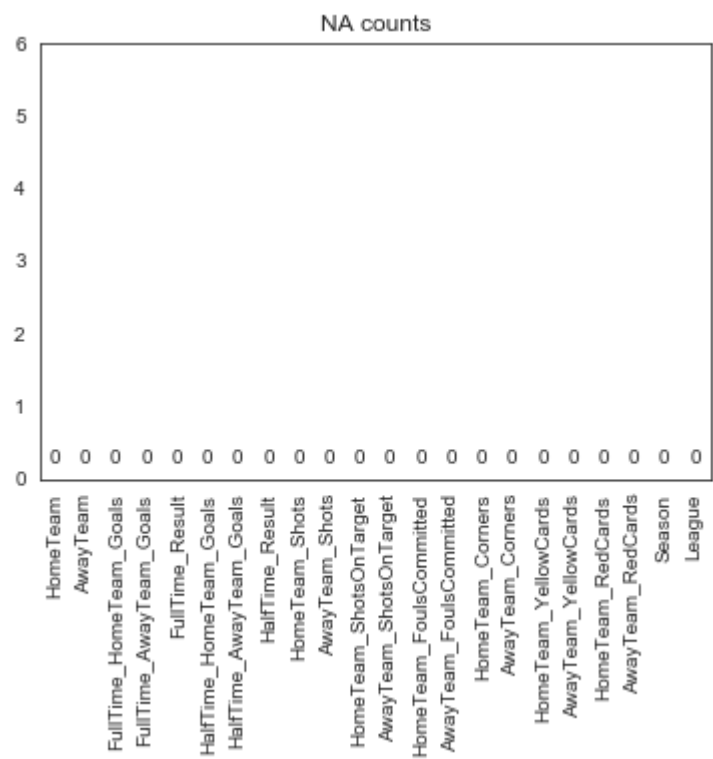
| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime |
|----------|----------|----------|-------------------------|-------------------------|----------|
| Date | | | | | |
| 23/09/12 | Cagliari | Roma | 0 | 3 | |
| 16/04/17 | Bastia | Lyon | 0 | 3 | |

Corner information is missing for the following matches :

- 1. Cagliari vs Roma
- 2. Bastia vs Lyon

We can use other matches involving these teams to populate the missing data

Now there should not be any row/column with null values



As expected, there are no null values now. All the null values have been cleaned up

Feature Engineering

Now lets add the following extra columns :

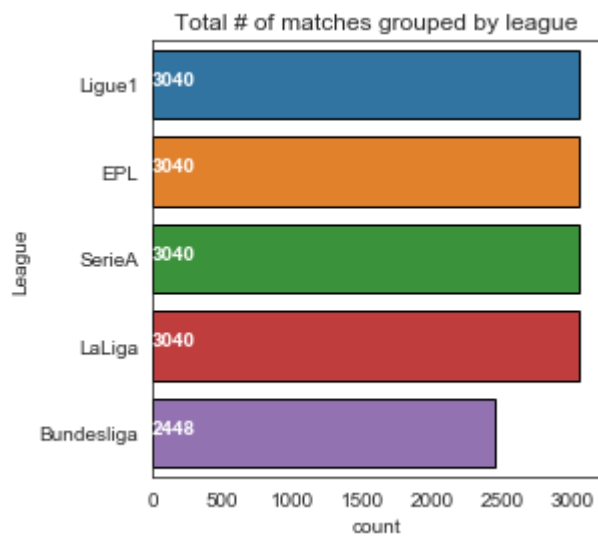
1. Winner : if FullTime_Result = 'H', then its Home Team. if FullTime_Result = 'A', then its Away Team. Else its null (implying draw)
2. Loser : if FullTime_Result = 'H', then its Home Team. if FullTime_Result = 'A', then its Away Team. Else its null (implying draw)
3. Total Cards (Home Team) : Home Team Yellow Cards + Home Team Red Cards
4. Total Cards (Away Team) : Away Team Yellow Cards + Away Team Red Cards
5. Percentage of Shots on Target (Home team): Home team shots on Target/Home team shots
6. Percentage of Shots on Target (Away team): Away team shots on Target/Away team shots
7. Home Team Goal saves = Away team shots on target - Full Time Away Team Goals
8. Away Team Goal saves = Home team shots on target - Full Time Home Team Goals
9. Total Goals = Full Time Home Team Goals + Full Time Away Team Goals
10. Total Cards = Total Cards (Home Team) + Total Cards(Away Team)

Let's look at the structure of our dataframe now

| | HomeTeam | AwayTeam | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | FullTime_R |
|------------|-------------|----------|-------------------------|-------------------------|------------|
| Date | | | | | |
| 2010-08-14 | Aston Villa | West Ham | 3 | 0 | |
| 2010-08-14 | Blackburn | Everton | 1 | 0 | |

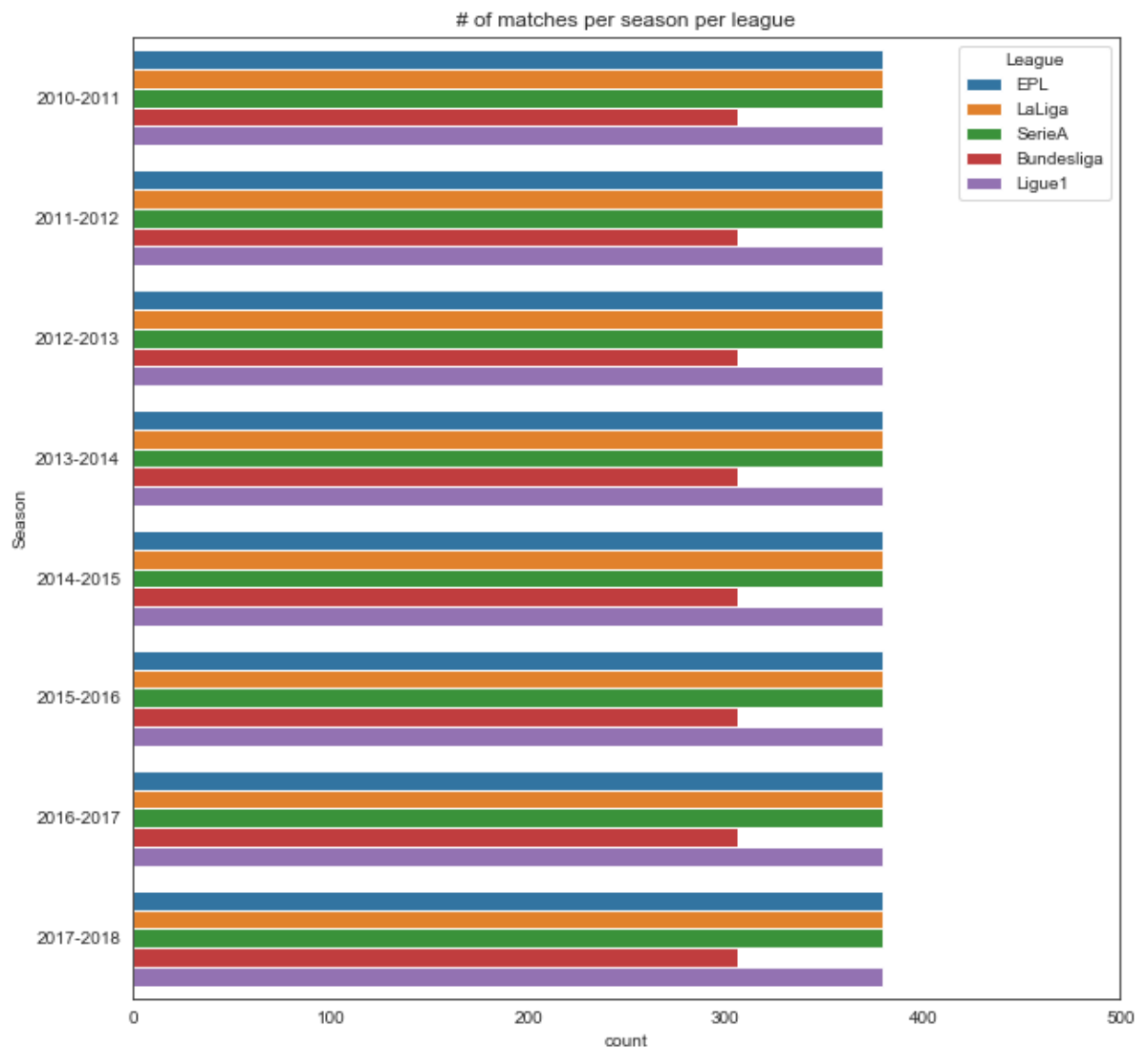
Lets have a look at the number of matches per season per league. If its not uniform, then we might need to adjust the aggregate data accordingly before comparing (We'll ignore the current season as not every league starts at the same time)

<Figure size 432x288 with 0 Axes>



Clearly Bundesliga has lesser number of matches as compared to other leagues. Lets see if number of matches per league varied across seasons

<Figure size 432x288 with 0 Axes>



So for a particular league, the number of matches per season has remained consistent over the years. So we can safely compare aggregate data for a league without adjusting (e.g total goals in EPL per season). However when we compare other league aggregate data with Bundesliga, the data might need some adjustment.

Bundesliga has 18 teams and 306 games (34 games for each team) per season whereas other leagues have 20 teams and 380 games a season (38 games for each team)

Hence when we compare data across leagues, aggregate data (such as total/count etc) for Bundesliga has to be scaled up accordingly in order to do a fair comparison

Feature Analysis

So finally we have the following features for analysis

HomeTeam
 AwayTeam
 FullTime_HomeTeam_Goals
 FullTime_AwayTeam_Goals
 FullTime_Result
 HalfTime_HomeTeam_Goals
 HalfTime_AwayTeam_Goals
 HalfTime_Result
 HomeTeam_Shots
 AwayTeam_Shots
 HomeTeam_ShotsOnTarget
 AwayTeam_ShotsOnTarget
 HomeTeam_FoulsCommitted
 AwayTeam_FoulsCommitted
 HomeTeam_Corners
 AwayTeam_Corners
 HomeTeam_YellowCards
 AwayTeam_YellowCards
 HomeTeam_RedCards
 AwayTeam_RedCards
 Season
 League
 Winner
 Loser
 TotalGoals
 HomeTeam_TotalCards
 AwayTeam_TotalCards
 TotalCards
 HomeTeam_ShotsOnTarget_Percent
 AwayTeam_ShotsOnTarget_Percent
 HomeTeam_GoalSaves
 AwayTeam_GoalSaves

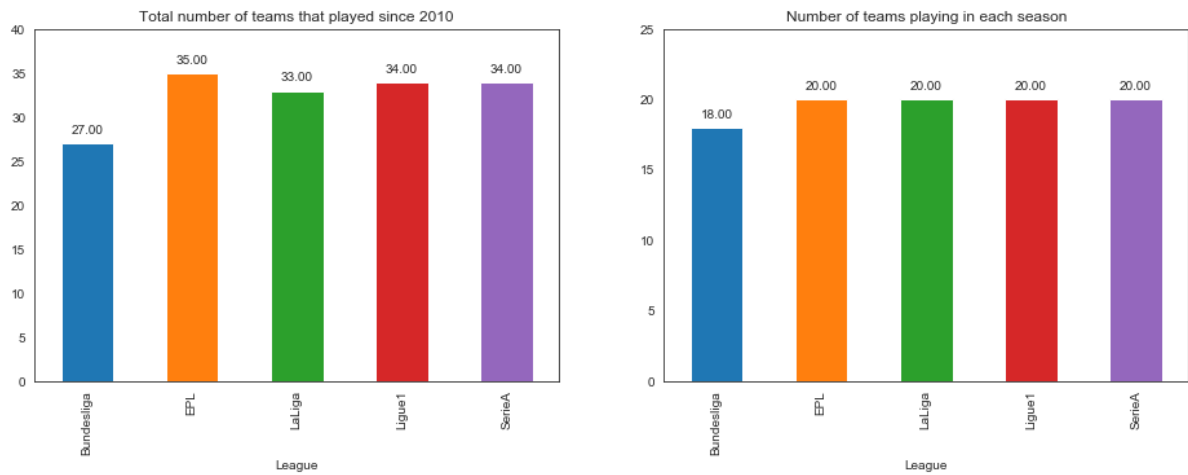
a) Home Team/Away Team : Contains the name of the home team/away team for a particular game

In every season, a team plays two games - one as home team another as away team. So the total number of unique values for HomeTeam and AwayTeam column should be identical

Home Team unique values # 163
 Away Team unique values # 163

Let's look at the total number of teams per league as well as the average number of teams per season

<Figure size 432x288 with 0 Axes>

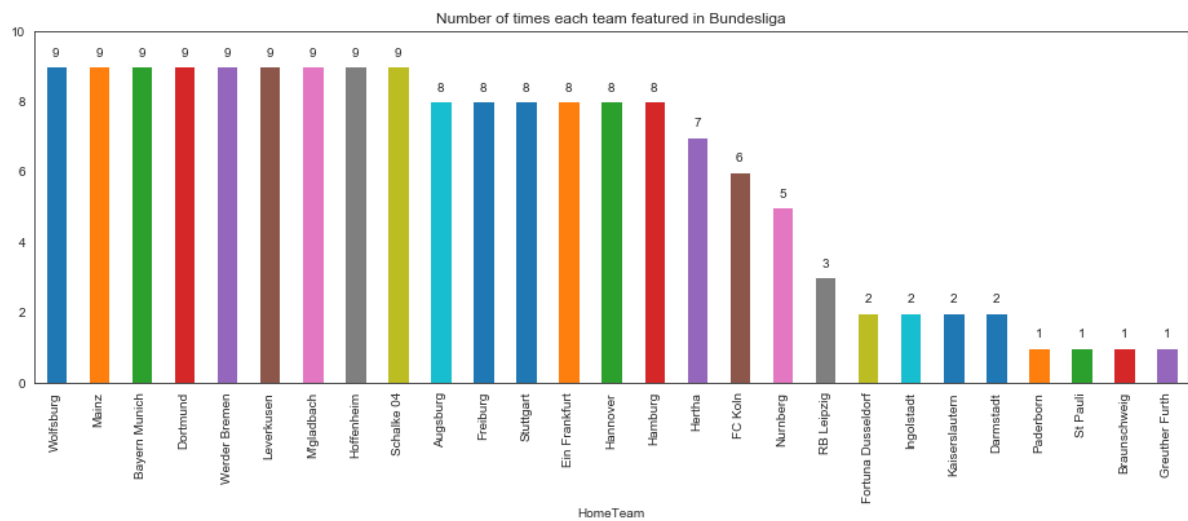


The difference between the two numbers indicates that there were occasions where teams were relegated to lower tier league or top teams from lower tier leagues were promoted

Let's look at the number of seasons each team featured in the league :

A. Bundesliga:

<Figure size 432x288 with 0 Axes>



The teams that have never been relegated from Bundesliga are :

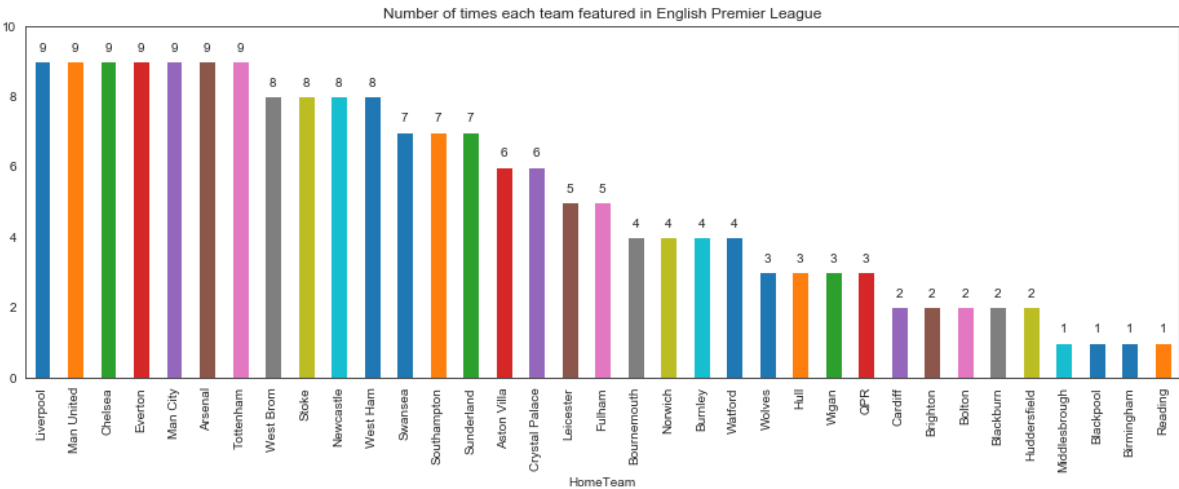
- 1. Wolfsburg
- 2. Mainz
- 3. Bayern Munich
- 4. Dortmund
- 5. Werder Bremen
- 6. Leverkusen
- 7. M'gladbach
- 8. Hoffenheim
- 9. Schalke04

Rest of the teams have been relegated at least once

9 out of 18 or 50 % of teams have remained consistent across last 9 seasons of Bundesliga

B. EPL

<Figure size 432x288 with 0 Axes>



Clearly the teams that have never been relegated from EPL are :

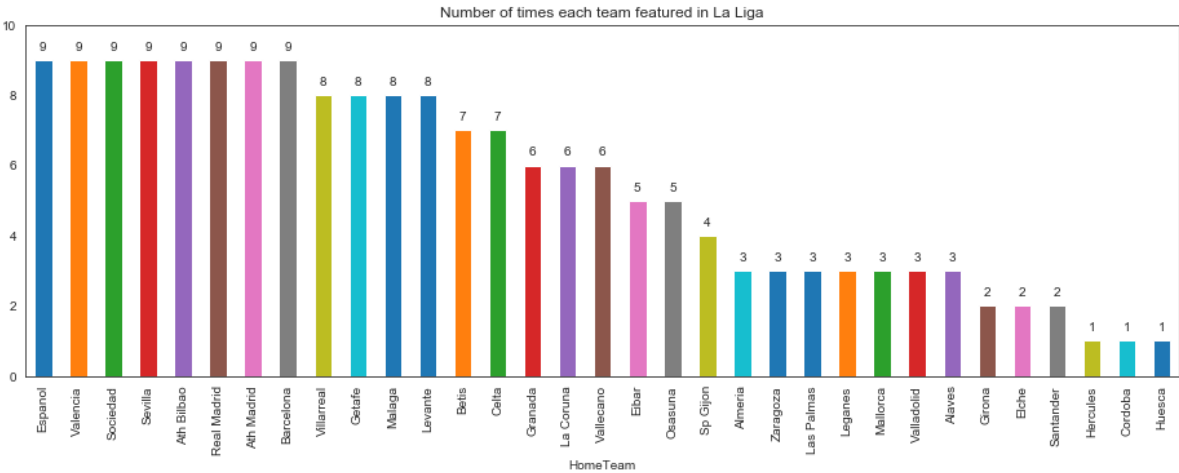
- 1. Liverpool
- 2. Man United
- 3. Chelsea
- 4. Everton
- 5. Man City
- 6. Arsenal
- 7. Tottenham

Rest of the teams have been relegated at least once

7 out of 20 or 35 % of teams have remained consistent across last 9 seasons of EPL

C. La Liga

<Figure size 432x288 with 0 Axes>



Clearly the teams that have never been relegated from LaLiga are :

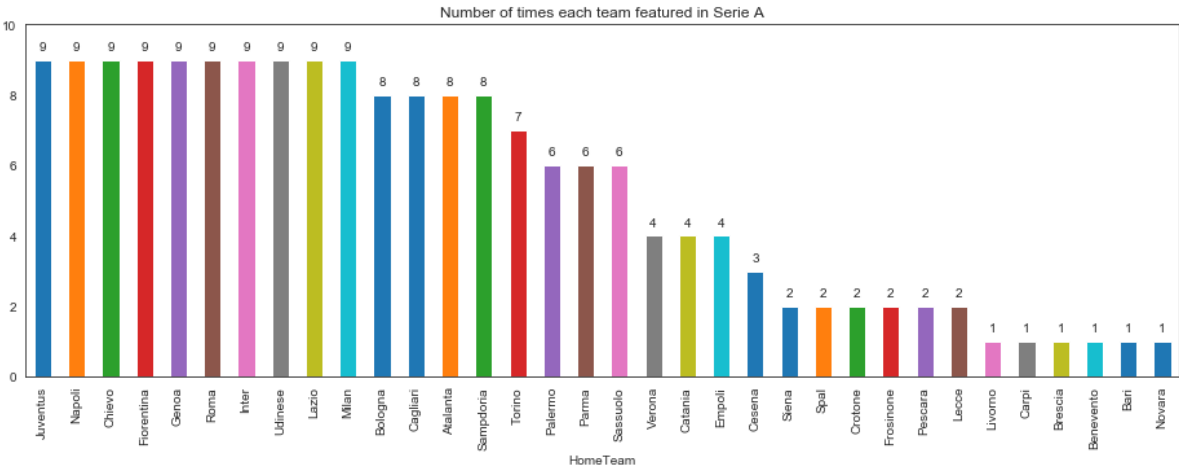
- 1. Espanol
- 2. Valencia
- 3. Sociedad
- 4. Sevilla
- 5. Ath Bilbao
- 6. Real Madrid
- 7. Ath Madrid
- 8. Barcelona

Rest of the teams have been relegated at least once

8 out of 20 or 40 % of teams have remained consistent across last 9 seasons of La Liga

D. Serie A

<Figure size 432x288 with 0 Axes>



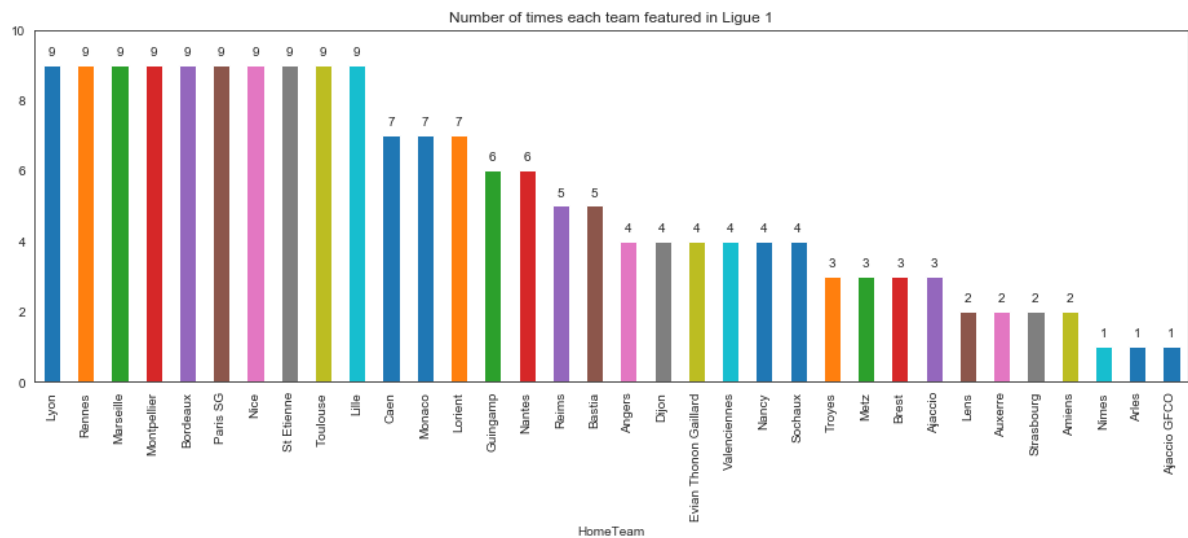
Clearly the teams that have never been relegated from Serie A are :

- 1. Juventus
- 2. Napoli
- 3. Chievo
- 4. Fiorentina
- 5. Genoa
- 6. Roma
- 7. Inter Milan
- 8. Udinese
- 9. Lazio
- 10. AC Milan

Rest of the teams have been relegated at least once
10 out of 20 or 50 % of teams have remained consistent across last 9 seasons of Serie A

E. Ligue 1

<Figure size 432x288 with 0 Axes>



Clearly the teams that have never been relegated from Ligue 1 are :

1. Lyon
2. Rennes
3. Marseille
4. Montpellier
5. Bordeaux
6. Paris SG
7. Nice
8. St Etienne
9. Toulouse
10. Lille

Rest of the teams have been relegated at least once

10 out of 20 or 50 % of teams have remained consistent across last 9 seasons of Serie A

In terms of relegation, EPL has been the most volatile league over past 9 years. Only 35 % teams have been consistent. For other leagues, the percentage varies from 40 to 50 %

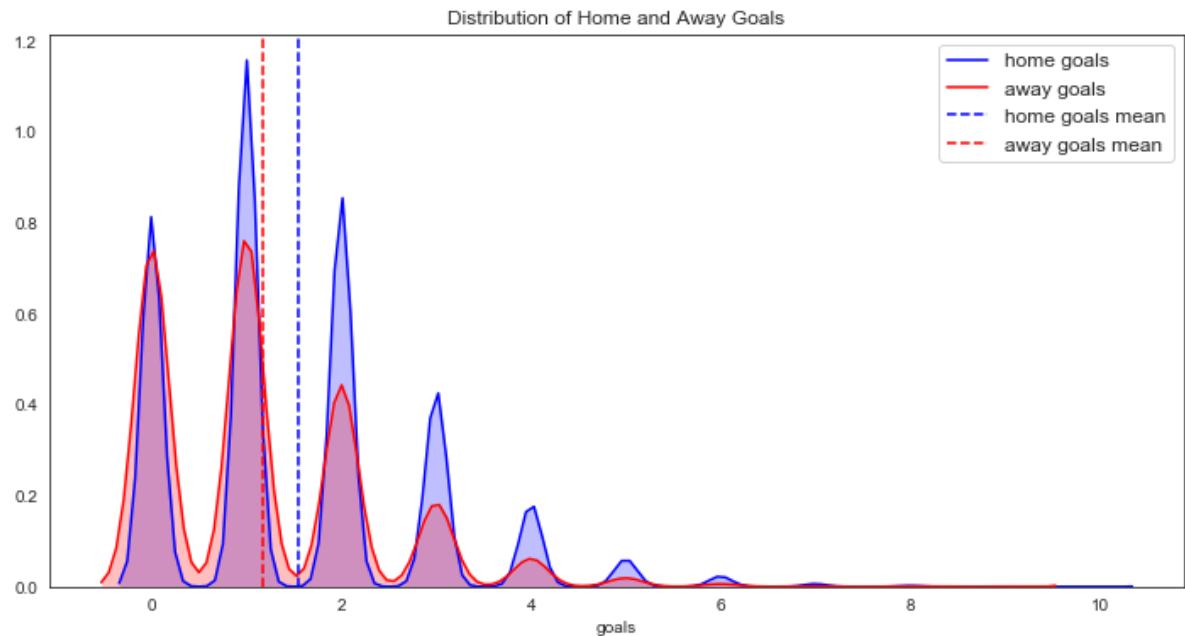
b. Full time Home/Away Team Goals/Total Goals : Indicates the number of goals scored by Home/Away team at the end of the match

Lets look at overall distribution of home and away goals first


```
C:\Users\ksaha\AppData\Local\Continuum\anaconda3\lib\site-packages\scipy\stats\stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.
```

```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```

<Figure size 432x288 with 0 Axes>



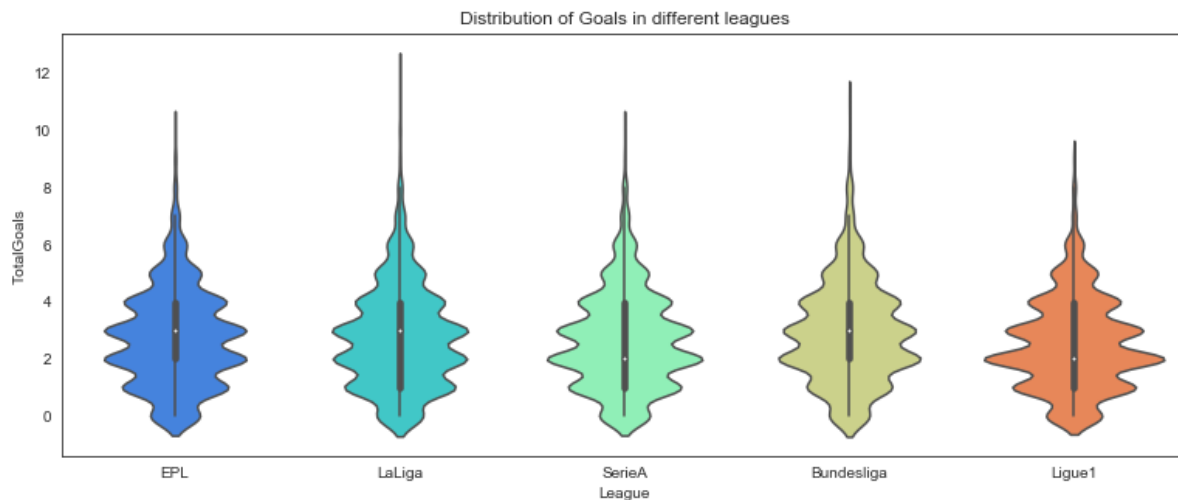
The mean value for home goals is higher than that of away goals. So there might be some home advantage.

Also lets take a look at how the distribution of total goals in a match varies across leagues

```
C:\Users\ksaha\AppData\Local\Continuum\anaconda3\lib\site-packages\scipy\stat
s\stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimension
al indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In th
e future this will be interpreted as an array index, `arr[np.array(seq)]`, wh
ich will result either in an error or a different result.
```

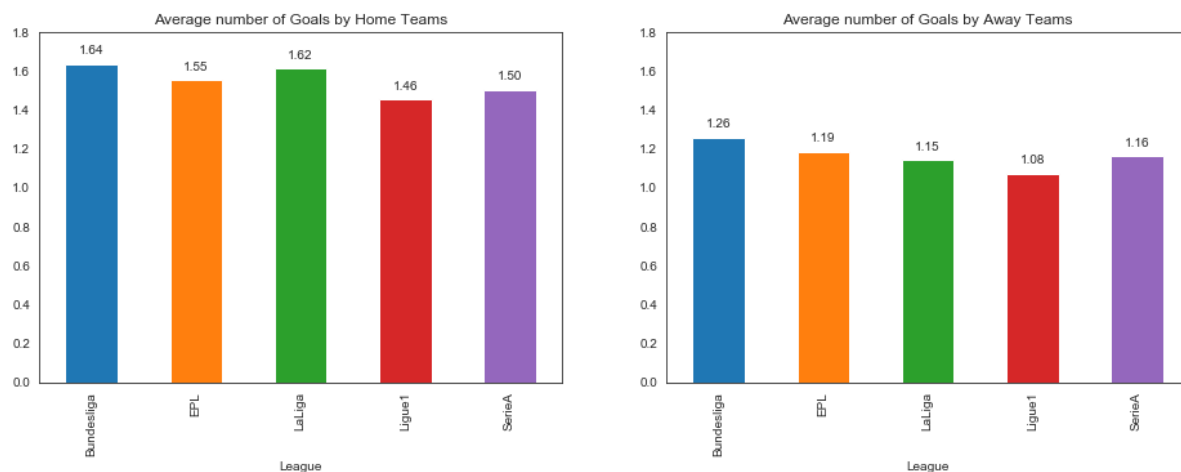
```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```

<Figure size 432x288 with 0 Axes>



Clearly for most of matches, the total number of goals varies between 1 and 4. It's quite rare to see more than 6 goals in a match in any league

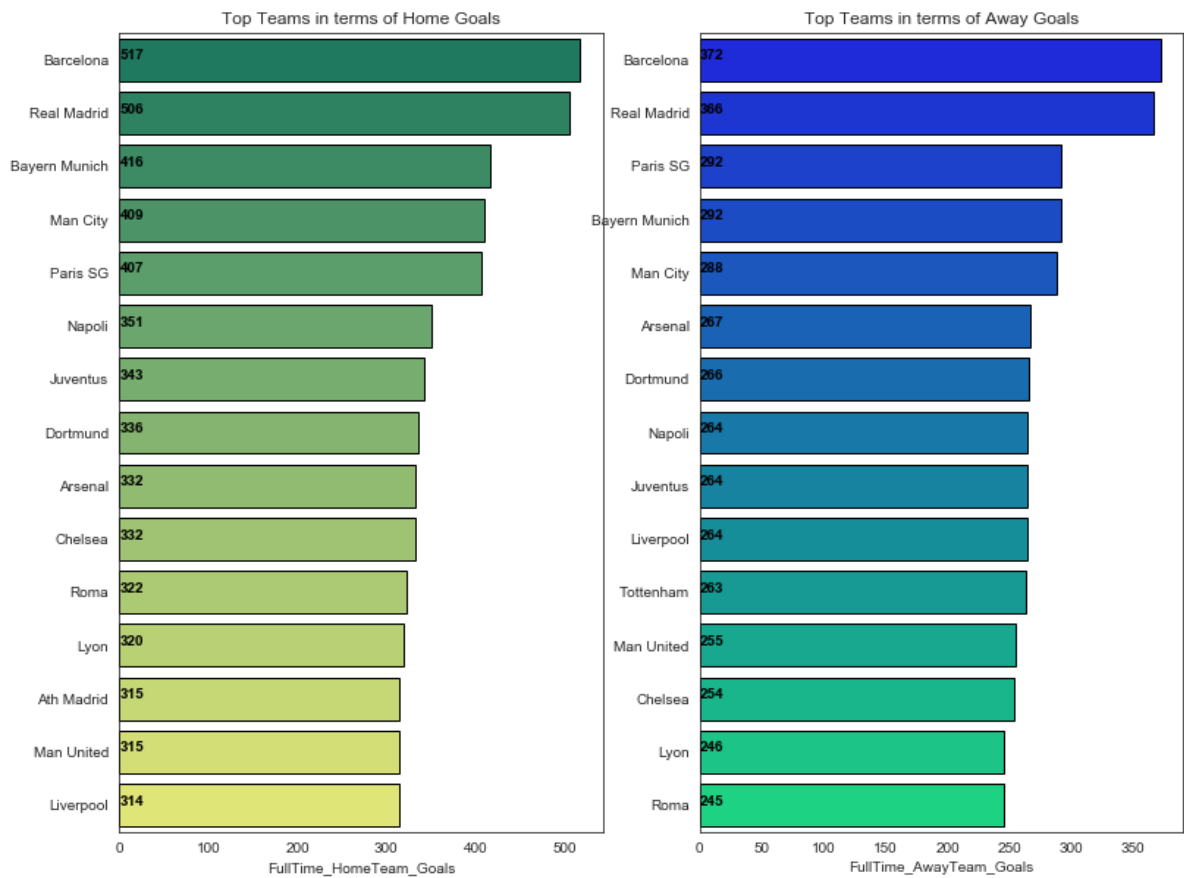
<Figure size 432x288 with 0 Axes>



Clearly in every league there's a significant home advantage

Lets look at the top teams in terms of number of home and away goals scored

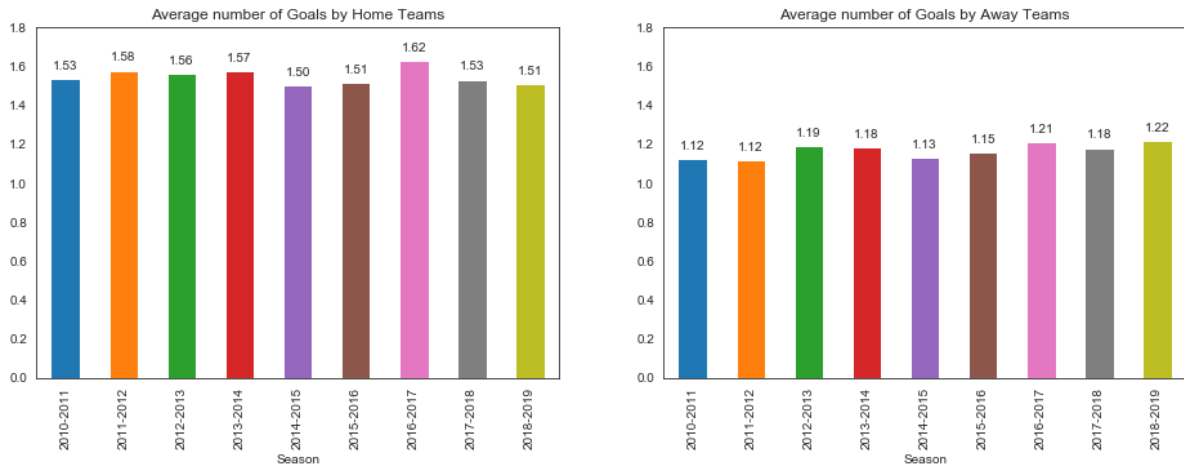
<Figure size 432x288 with 0 Axes>



These two lists have a lot of teams in common. This indicates that good teams usually do well at home as well as away (However they score better at home than at away).

Let's look at variation of goal data across last 9 seasons

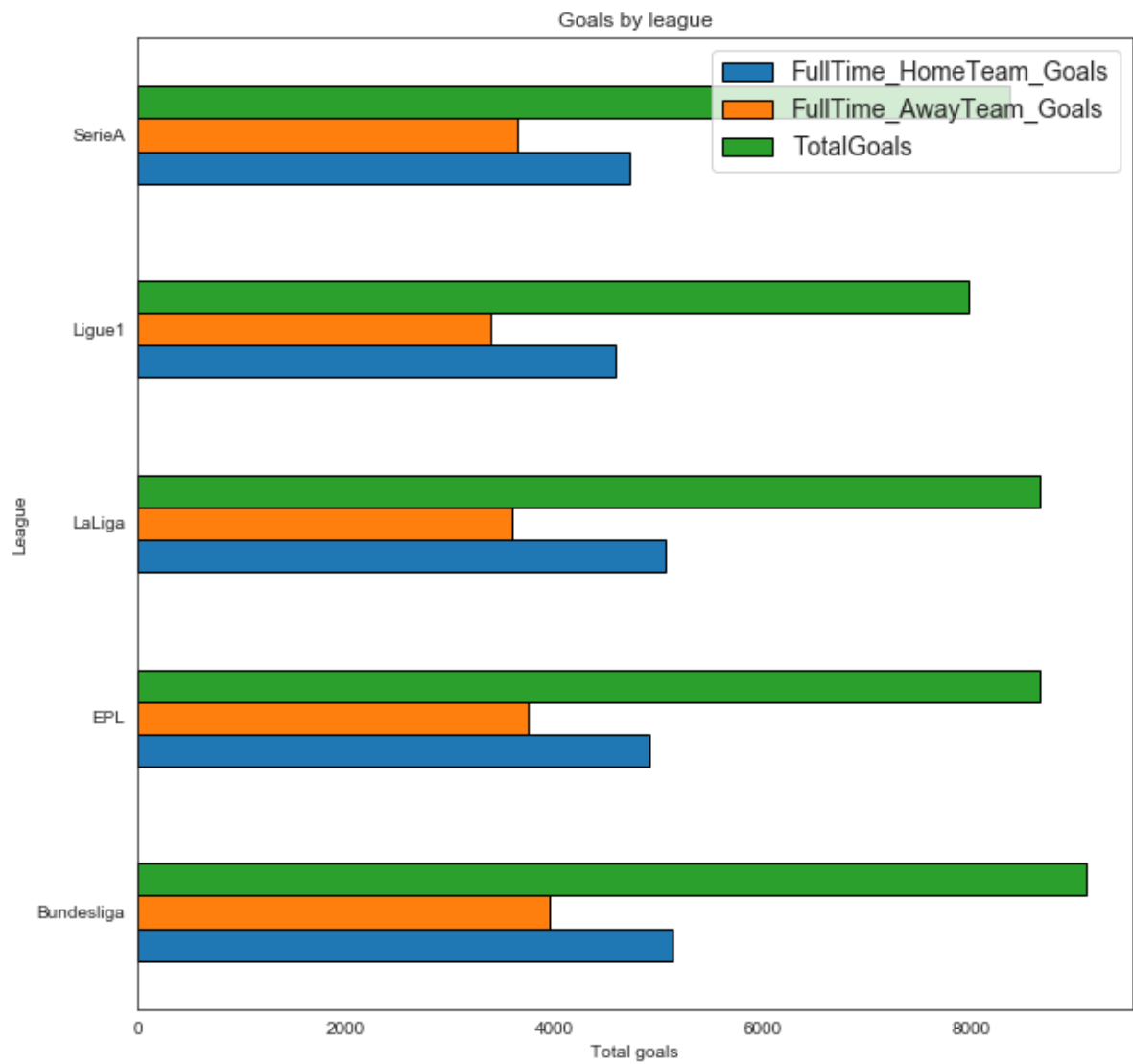
<Figure size 432x288 with 0 Axes>



In last 10 years, overall, away teams have improved (but only marginally) - from 1.12 goals per match to 1.22 goals per match

Lets look at the total number of goals as well. Let’s find out which league had the most goals. To make it a fair comparison we will adjust Bundesliga numbers (Bundesliga has 18 teams as opposed to 20 for other leagues)

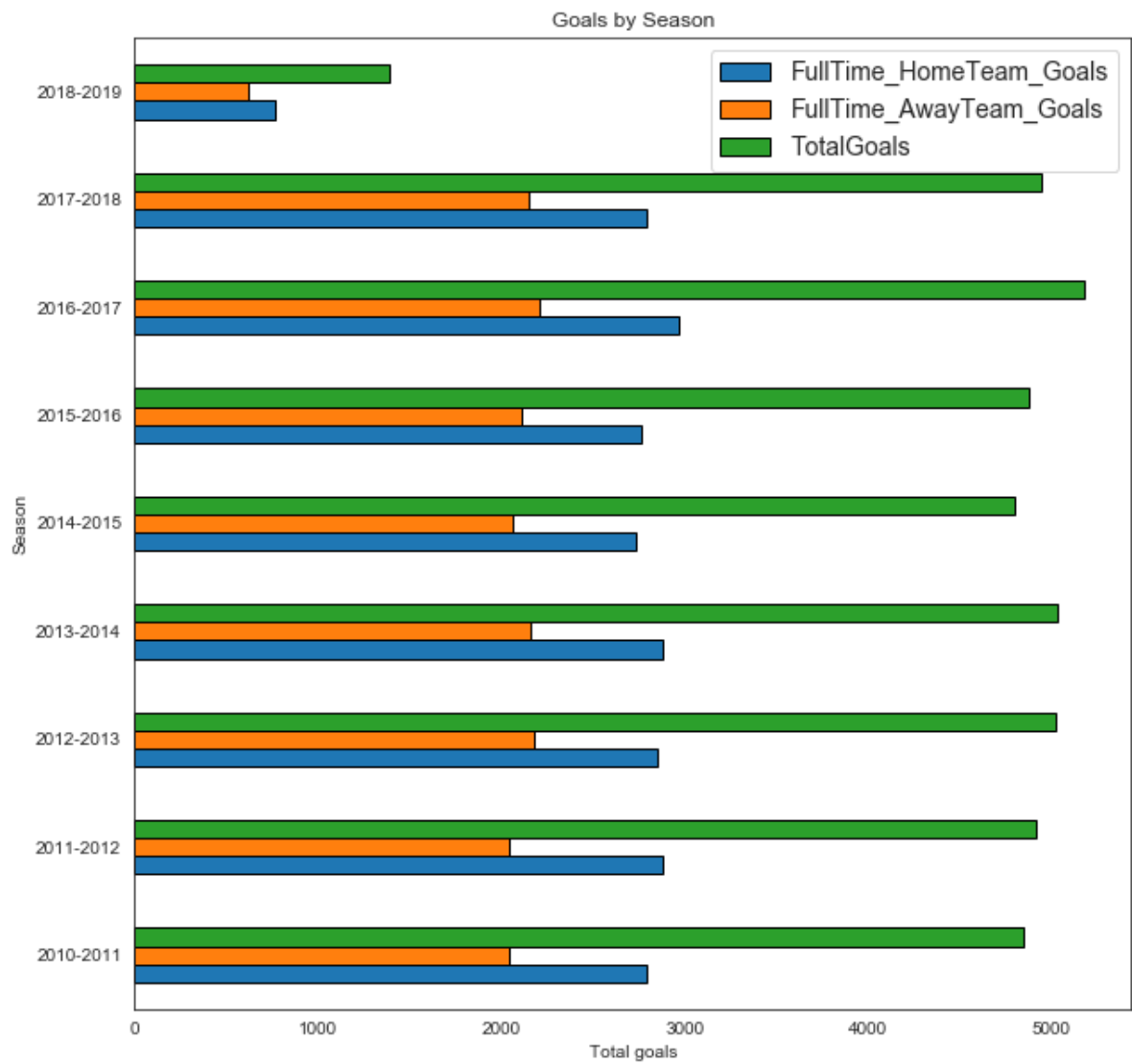
<Figure size 432x288 with 0 Axes>



Clearly Ligue One produces less goals as compared to other leagues. Post adjustment, Bundesliga produces more goals than anybody else.

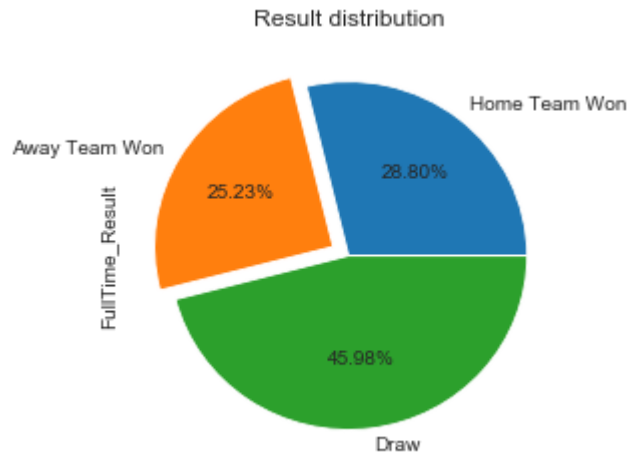
Lets look at the season-wise data as well.

<Figure size 432x288 with 0 Axes>



From 2010-11 to 2013-14, number of goals increased each season but stopped growing after 2013-14 season (except for a spike in 2016-17)

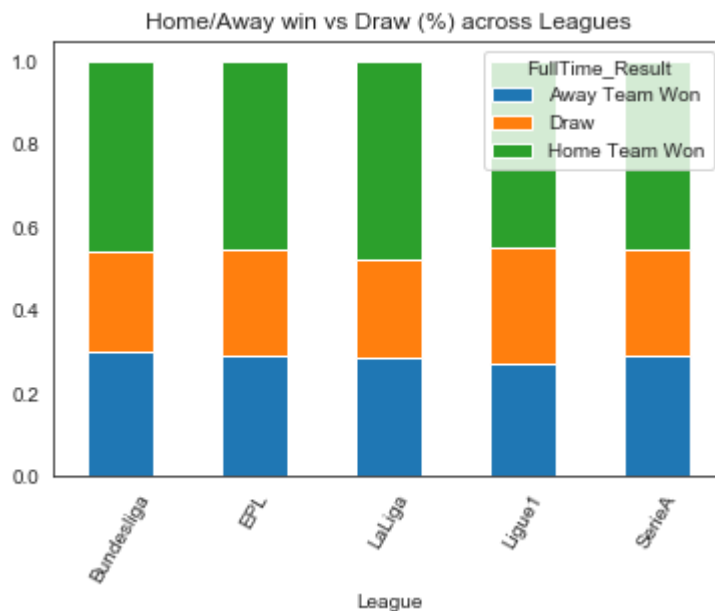
c. Full Time Result : Indicates whether Home or Away team won or if the match ended in a draw



Overall, home team wins nearly 46 % of matches

Lets look at distribution of match results in various league

<Figure size 432x288 with 0 Axes>



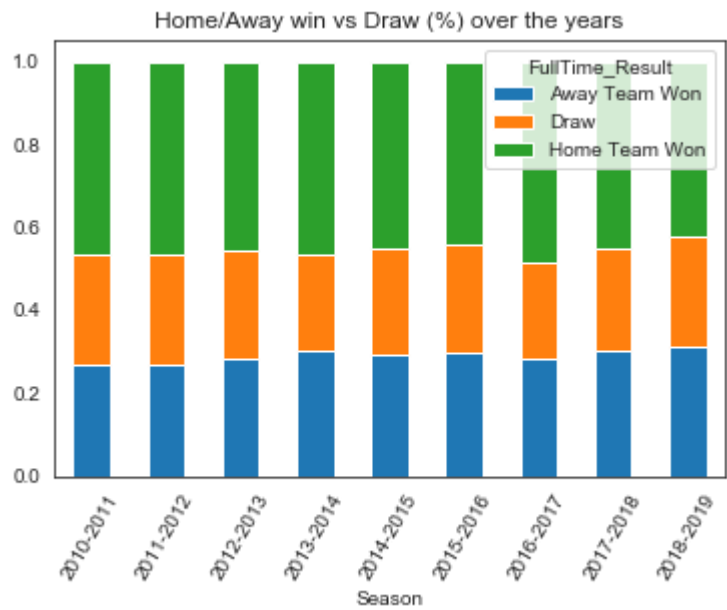
Home Team wins more than 40 % of matches in every league.

Home Advantage seems to be a bigger factor in La Liga as compared to other leagues

Ligue1 produces more draws than other leagues

Lets look at how the data changed over the years

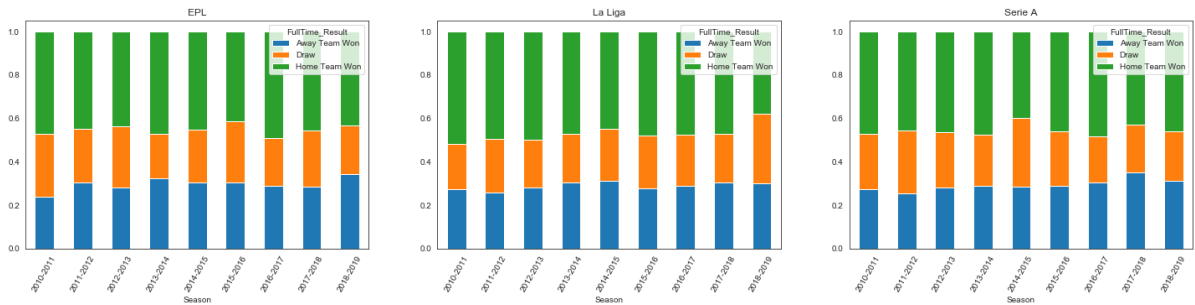
<Figure size 432x288 with 0 Axes>



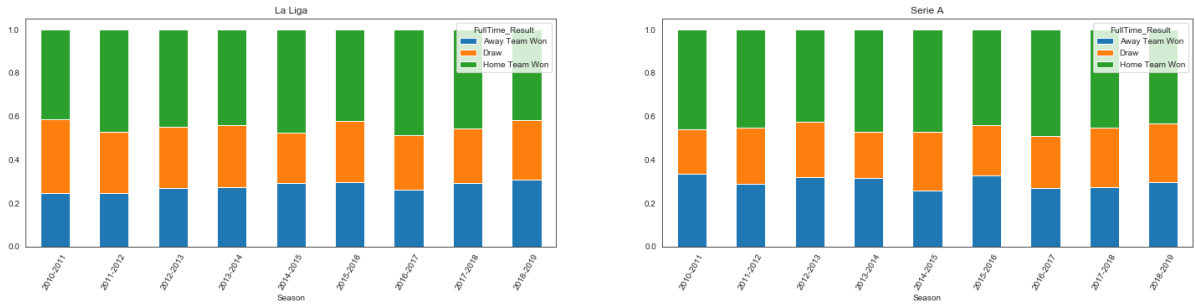
There hasn't been any drastic change over the years - Home advantage has always been significant. Away performances have improved but only marginally

Lets look at how the data distribution varied from league-to-league over last 9 years

<Figure size 432x288 with 0 Axes>



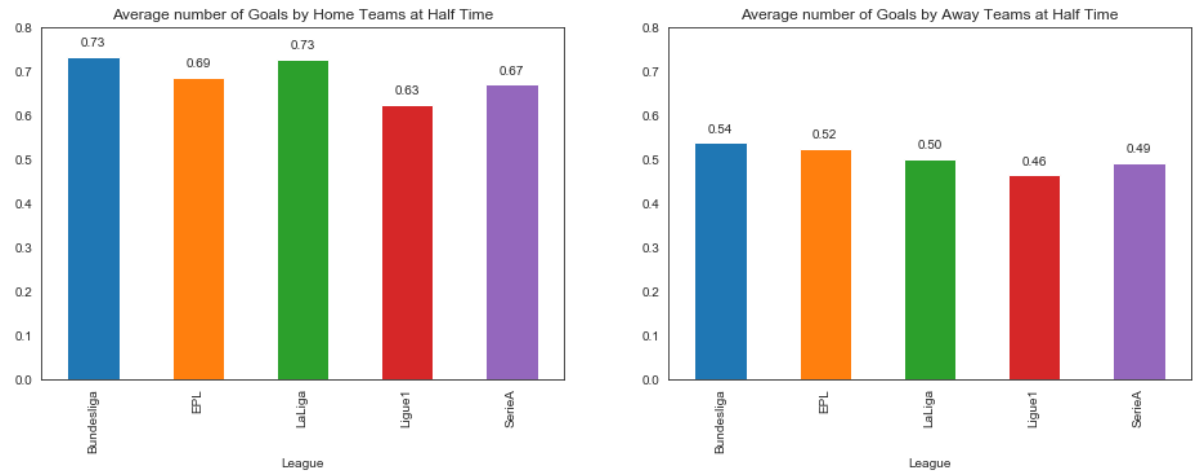
<Figure size 432x288 with 0 Axes>



There is no league specific trend that emerged over the years. Home Advantage continues to be significant across leagues across seasons

d. Half Time Home/Away team goals : Indicates the number of Goals scored by Home/Away team at Half Time

<Figure size 432x288 with 0 Axes>



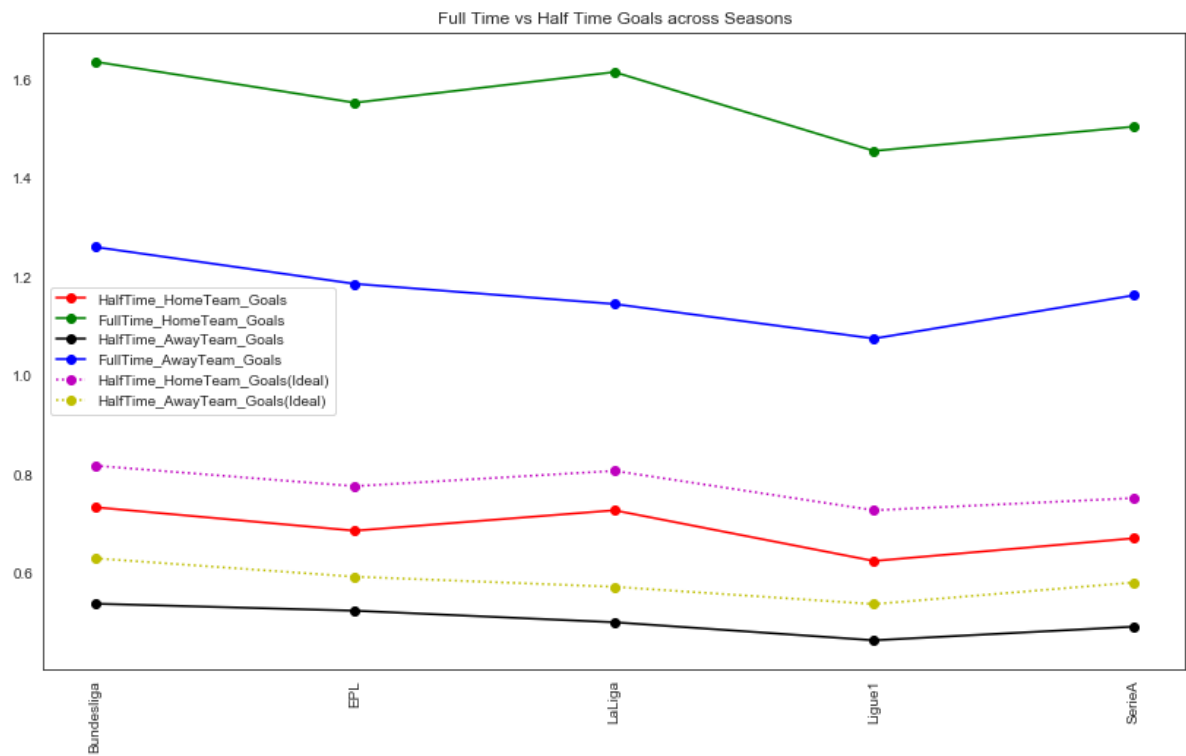
So Home Team typically dominates right from the beginning

If we look at the corresponding full time goal numbers :

Out[83]:

| League | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | HalfTime_HomeTeam_Goals | H |
|------------|-------------------------|-------------------------|-------------------------|---|
| Bundesliga | 1.636220 | 1.260973 | 0.733887 | |
| EPL | 1.553165 | 1.186392 | 0.686392 | |
| LaLiga | 1.615287 | 1.145541 | 0.727707 | |
| Ligue1 | 1.455556 | 1.075556 | 0.625079 | |
| SerieA | 1.504777 | 1.163057 | 0.671019 | |

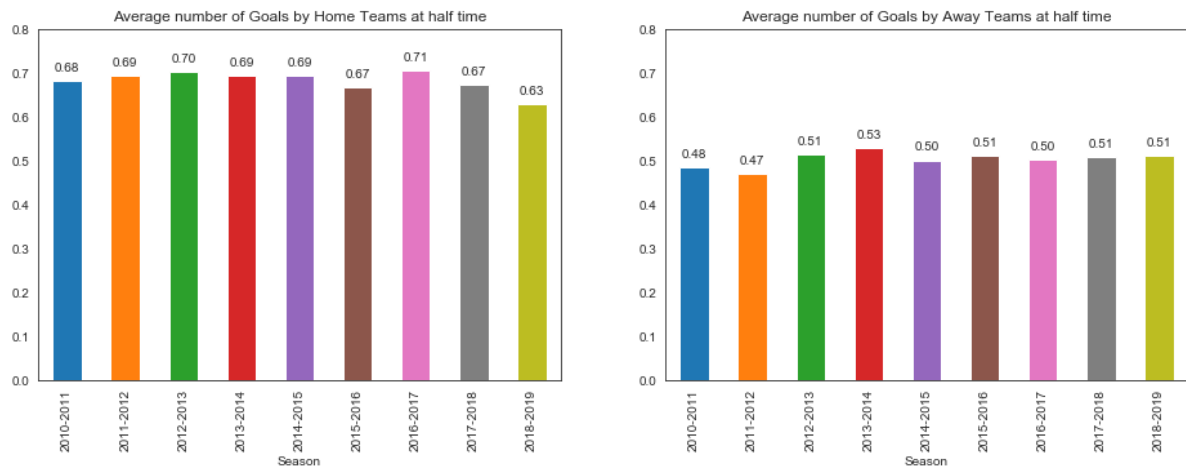
<Figure size 432x288 with 0 Axes>



For Both Home and Away teams, Half Time goals is significantly less than full time goals/2

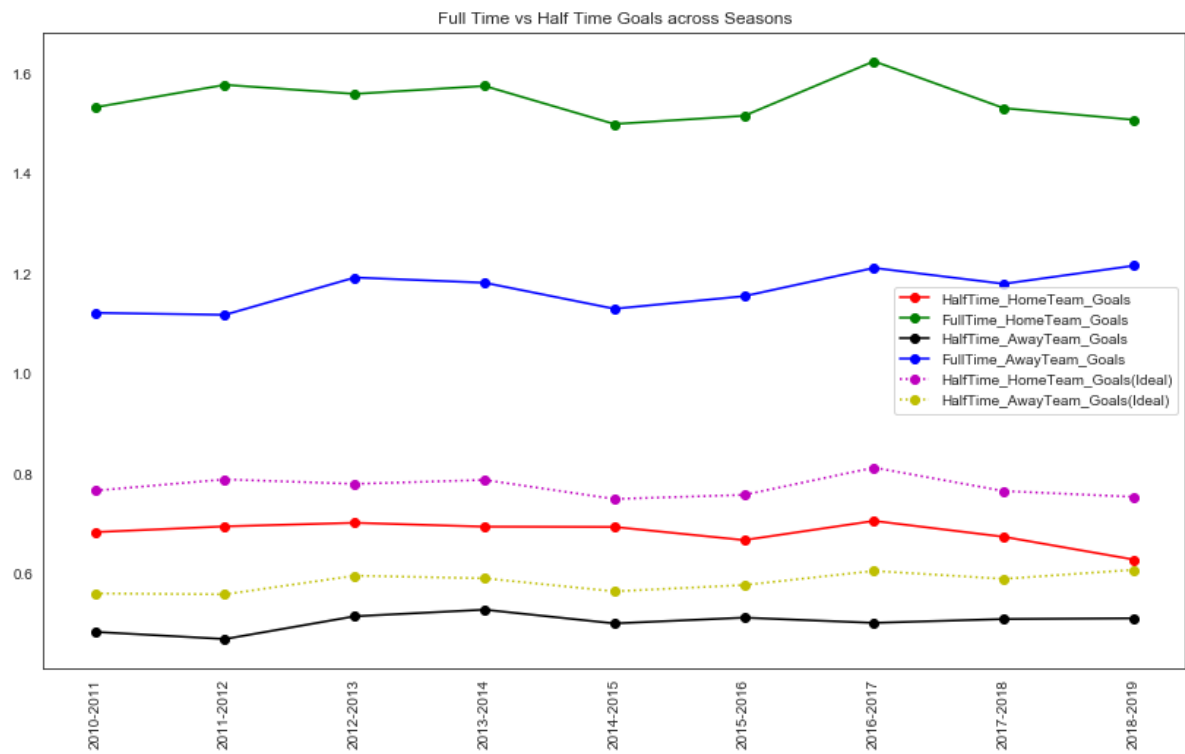
So both Home and Away teams are typically more aggressive and attack more during second half of the match

<Figure size 432x288 with 0 Axes>



Data pattern remains almost uniform across years

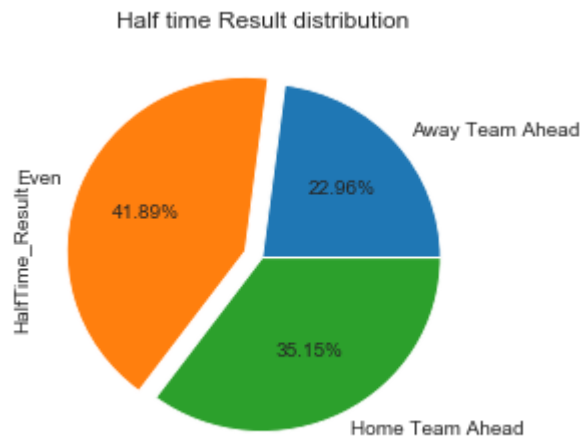
<Figure size 432x288 with 0 Axes>



Over the years, both home and away teams have remained defensive during first half and more aggressive during second half

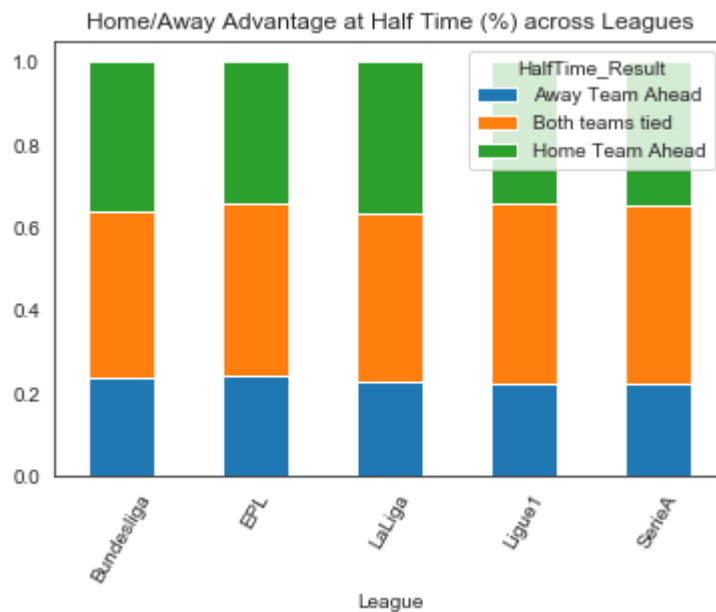
e. Half Time Result : Indicates whether Home or Away team was ahead at half time (or if both teams were tied at equal no of goals)

Lets look at overall distribution of data first



Home Team advantage falls but only by an insignificant margin. Let's look at how the data varied between different leagues

<Figure size 432x288 with 0 Axes>



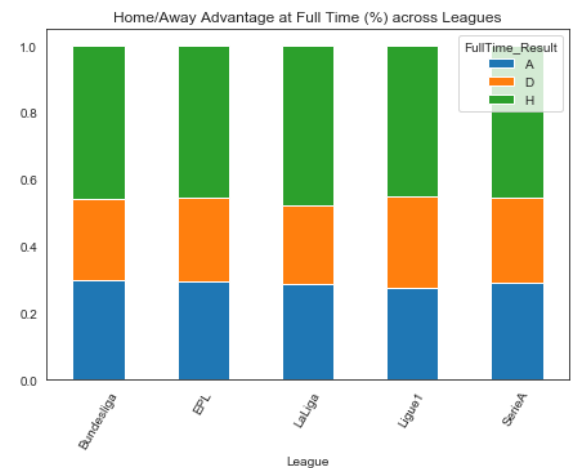
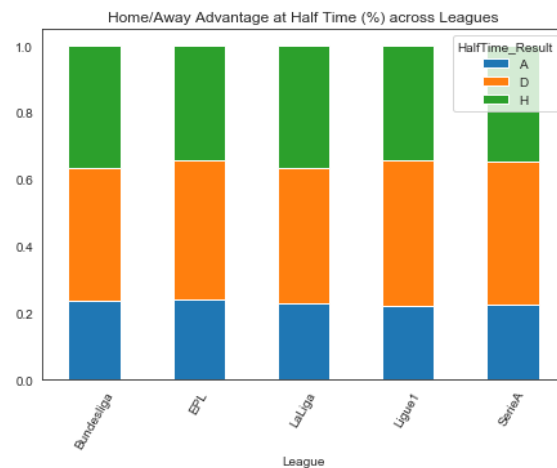
Home Team is ahead at half time in nearly 40 % of matches in every league. So home team dominates right from beginning

Lets look at the half time and full time data together

| HalfTime_Result | A | D | H |
|-------------------|----------|----------|----------|
| League | | | |
| Bundesliga | 0.237248 | 0.398577 | 0.364176 |
| EPL | 0.239557 | 0.417722 | 0.342722 |
| LaLiga | 0.228662 | 0.404777 | 0.366561 |
| Ligue1 | 0.220952 | 0.438095 | 0.340952 |
| SerieA | 0.222930 | 0.431210 | 0.345860 |

| FullTime_Result | A | D | H |
|-------------------|----------|----------|----------|
| League | | | |
| Bundesliga | 0.298142 | 0.243575 | 0.458284 |
| EPL | 0.292405 | 0.253165 | 0.454430 |
| LaLiga | 0.287261 | 0.233439 | 0.479299 |
| Ligue1 | 0.273016 | 0.276190 | 0.450794 |
| SerieA | 0.291083 | 0.253185 | 0.455732 |

<Figure size 432x288 with 0 Axes>



From the data the following points are clear :

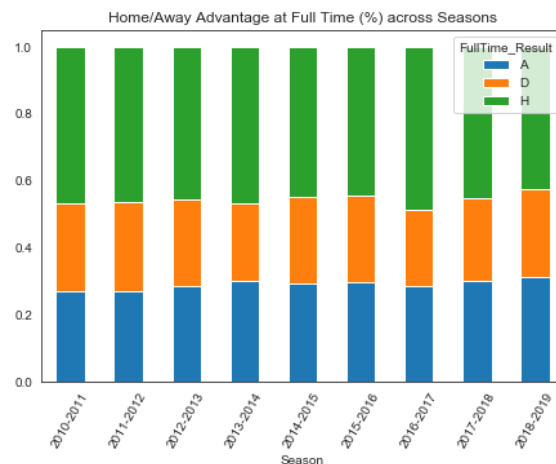
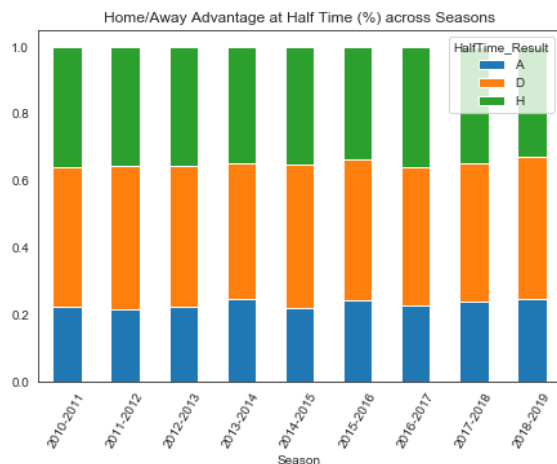
1. Teams are more aggressive in second half and push for a result. Nearly 40 % of matches are tied at half time while only 23-26 % matches are tied at full time
2. Home Advantage is stronger in second half than in the first half

Lets look at how the data changed over the years

| HalfTime_Result | A | D | H |
|-----------------|----------|----------|----------|
| Season | | | |
| 2010-2011 | 0.224535 | 0.415663 | 0.359803 |
| 2011-2012 | 0.215772 | 0.427163 | 0.357065 |
| 2012-2013 | 0.223987 | 0.419496 | 0.356517 |
| 2013-2014 | 0.246988 | 0.403067 | 0.349945 |
| 2014-2015 | 0.216867 | 0.429901 | 0.353231 |
| 2015-2016 | 0.241512 | 0.422234 | 0.336254 |
| 2016-2017 | 0.225082 | 0.416210 | 0.358708 |
| 2017-2018 | 0.237130 | 0.416210 | 0.346659 |
| 2018-2019 | 0.246575 | 0.422701 | 0.330724 |

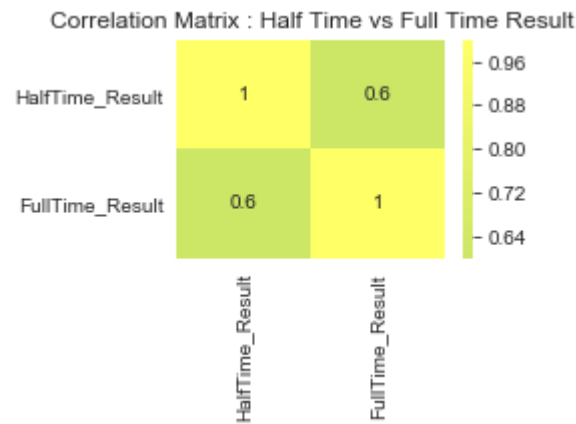
| FullTime_Result | A | D | H |
|-----------------|----------|----------|----------|
| Season | | | |
| 2010-2011 | 0.269989 | 0.262870 | 0.467141 |
| 2011-2012 | 0.268894 | 0.265608 | 0.465498 |
| 2012-2013 | 0.284775 | 0.259584 | 0.455641 |
| 2013-2014 | 0.300110 | 0.233297 | 0.466594 |
| 2014-2015 | 0.290800 | 0.259584 | 0.449617 |
| 2015-2016 | 0.297371 | 0.259036 | 0.443593 |
| 2016-2017 | 0.283133 | 0.230559 | 0.486309 |
| 2017-2018 | 0.301752 | 0.244797 | 0.453450 |
| 2018-2019 | 0.313112 | 0.262231 | 0.424658 |

<Figure size 432x288 with 0 Axes>



Lets calculate the correlation coefficient between half time and full time results

<Figure size 432x288 with 0 Axes>



There is some coorelation between half time and full time results but not exactly a linear relationship

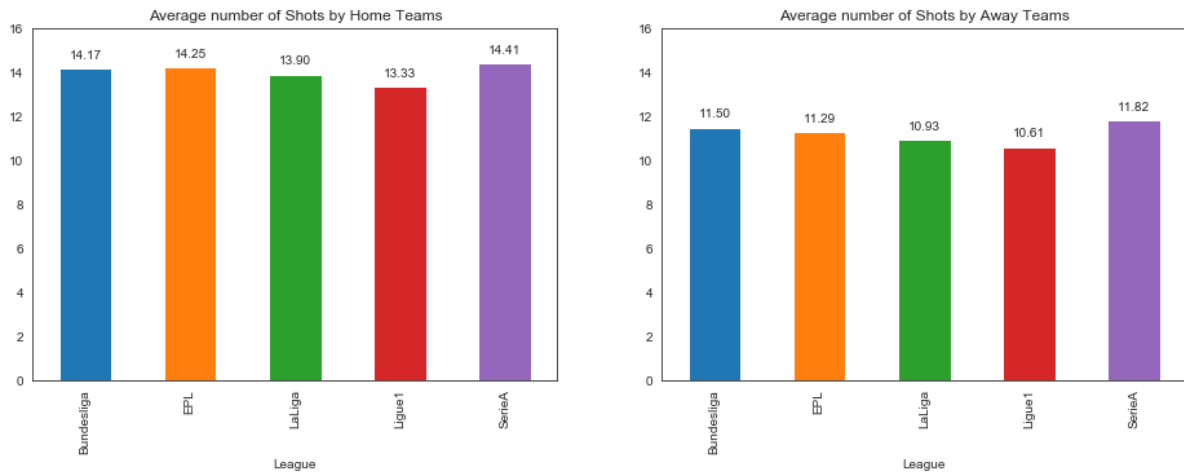
If we calculate correlation using Spearman approach instead of Pearson approach, we get similar numbers as well

0.5944584330392942

```
C:\Users\ksaha\AppData\Local\Continuum\anaconda3\lib\site-packages\scipy\stats\stats.py:245: RuntimeWarning: The input array could not be properly checked for nan values. nan values will be ignored.  
"values. nan values will be ignored.", RuntimeWarning)
```

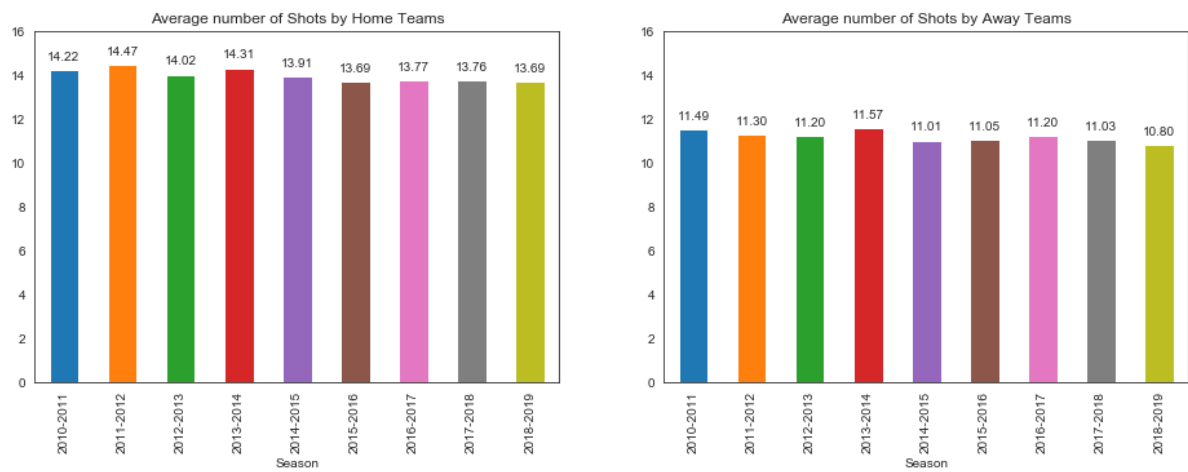
f. Home/Away Team Shots: This represents the number of shots at goal attempted by home or away team during a match

<Figure size 432x288 with 0 Axes>



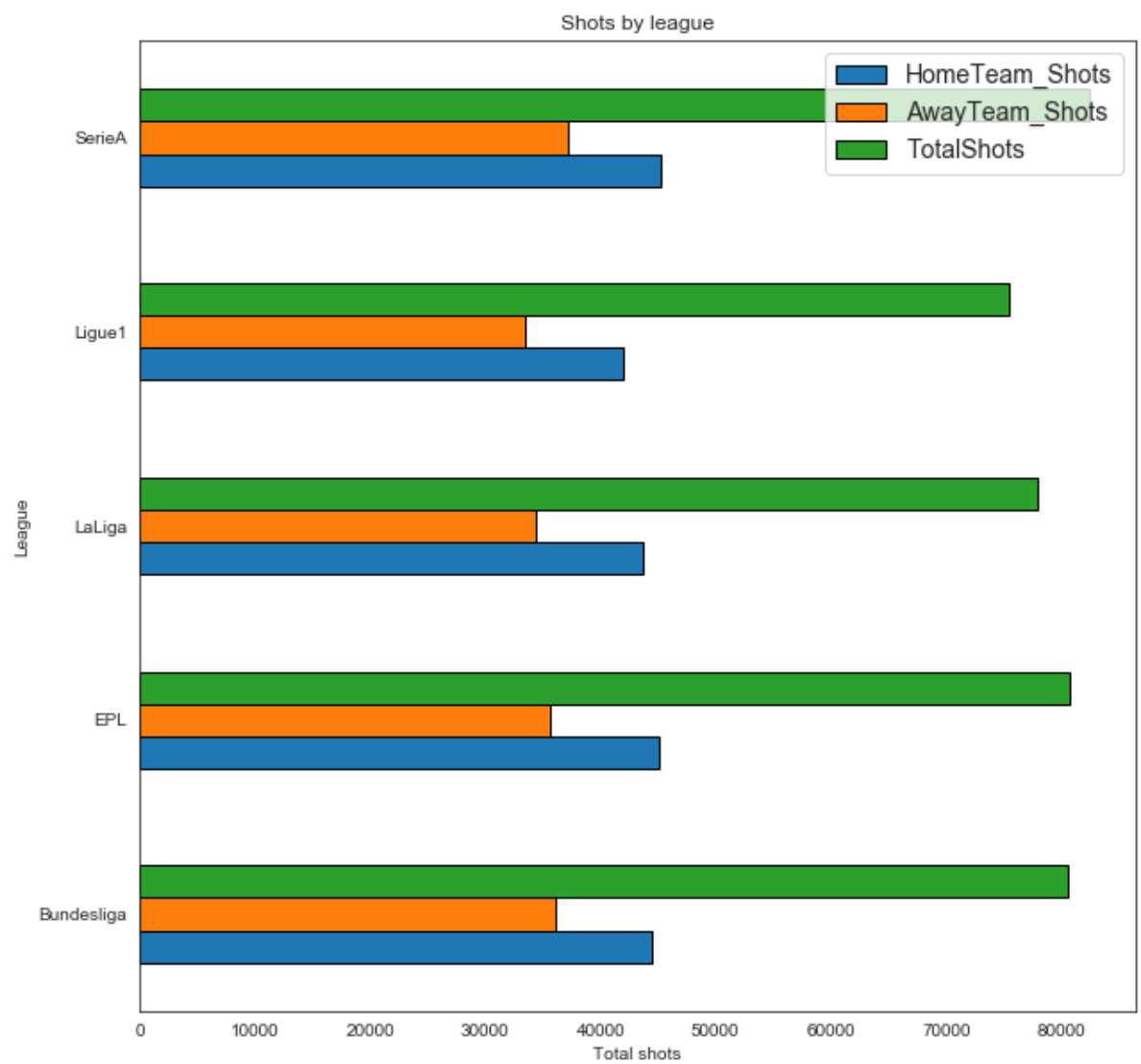
Home Teams are ahead here as well. Lets look at the trend over last 10 seasons

<Figure size 432x288 with 0 Axes>



Lets look at the total number of shots as well. Total Number of shots = home team shots + away team shots

<Figure size 432x288 with 0 Axes>



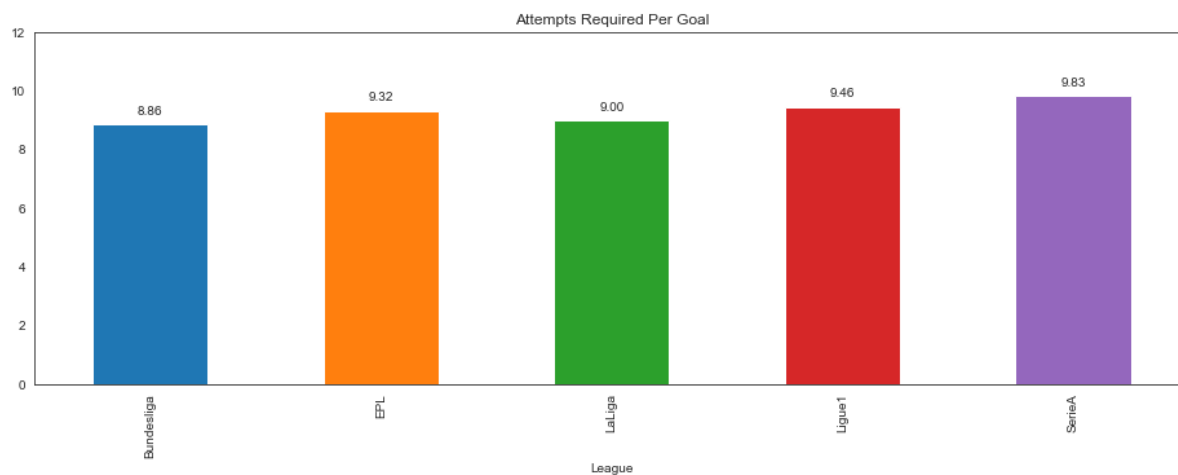
Serie A, Bundesliga (adjusted) and EPL has max number of shots at goal per season.

Interestingly, La Liga leads in number of goals per season but not at number of shots whereas it is other way round for Serie A

So, we can conclude that La Liga forwards are more accurate in general as compared to Serie A forwards

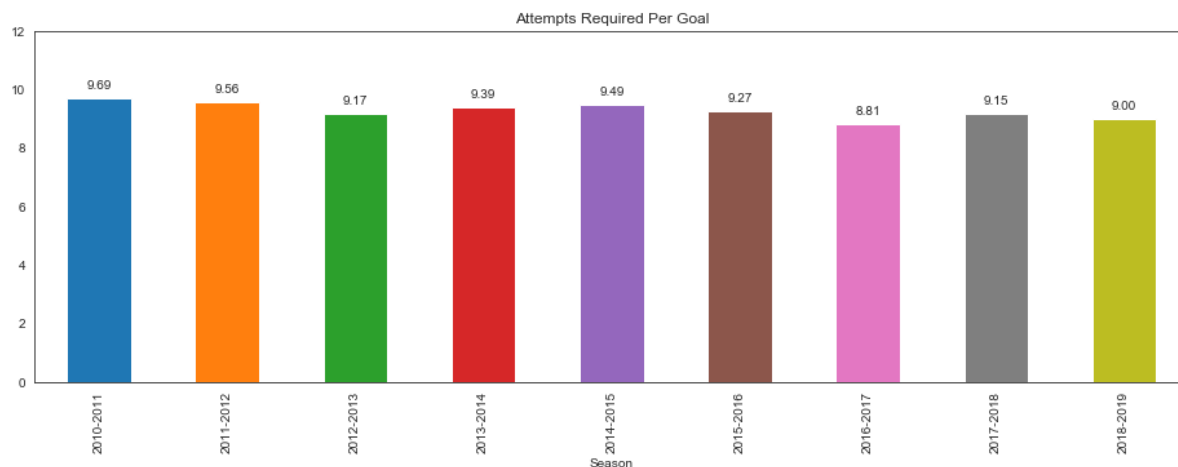
Now that we have both goal and shots data,. A good measure of accuracy will be number of attempts required to score a goal successfully we can take a look at accuracy

<Figure size 432x288 with 0 Axes>



Forwards in Bundesliga and LaLiga are more accurate as compared to others. Forwards in SerieA are the least effective

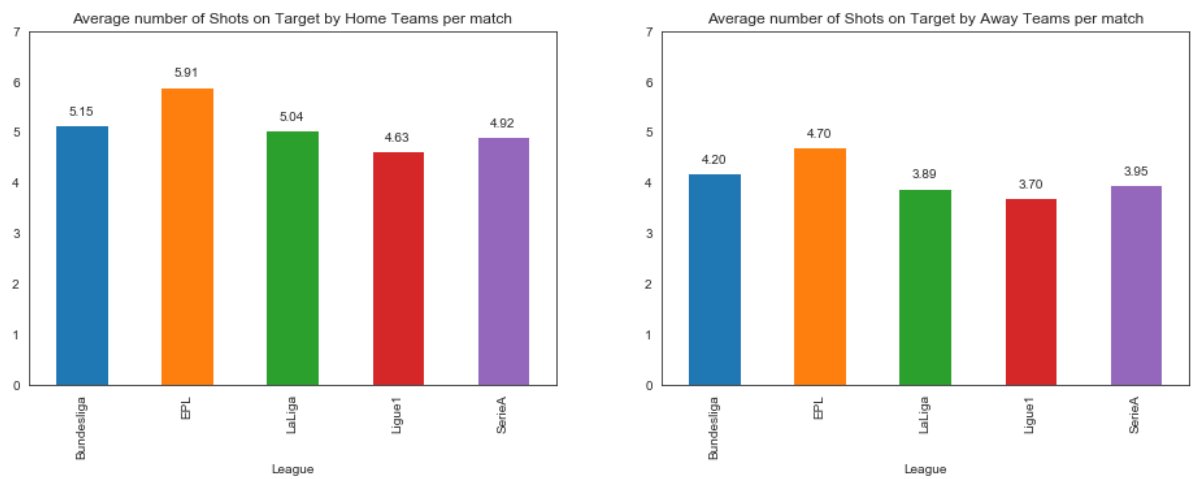
<Figure size 432x288 with 0 Axes>



Over the years, forwards are getting more effective. But overall change is marginal

g. Home/Away team shots on target : This captures the number of shots on target (scored goals or hit the bar or saved by goal keeper)

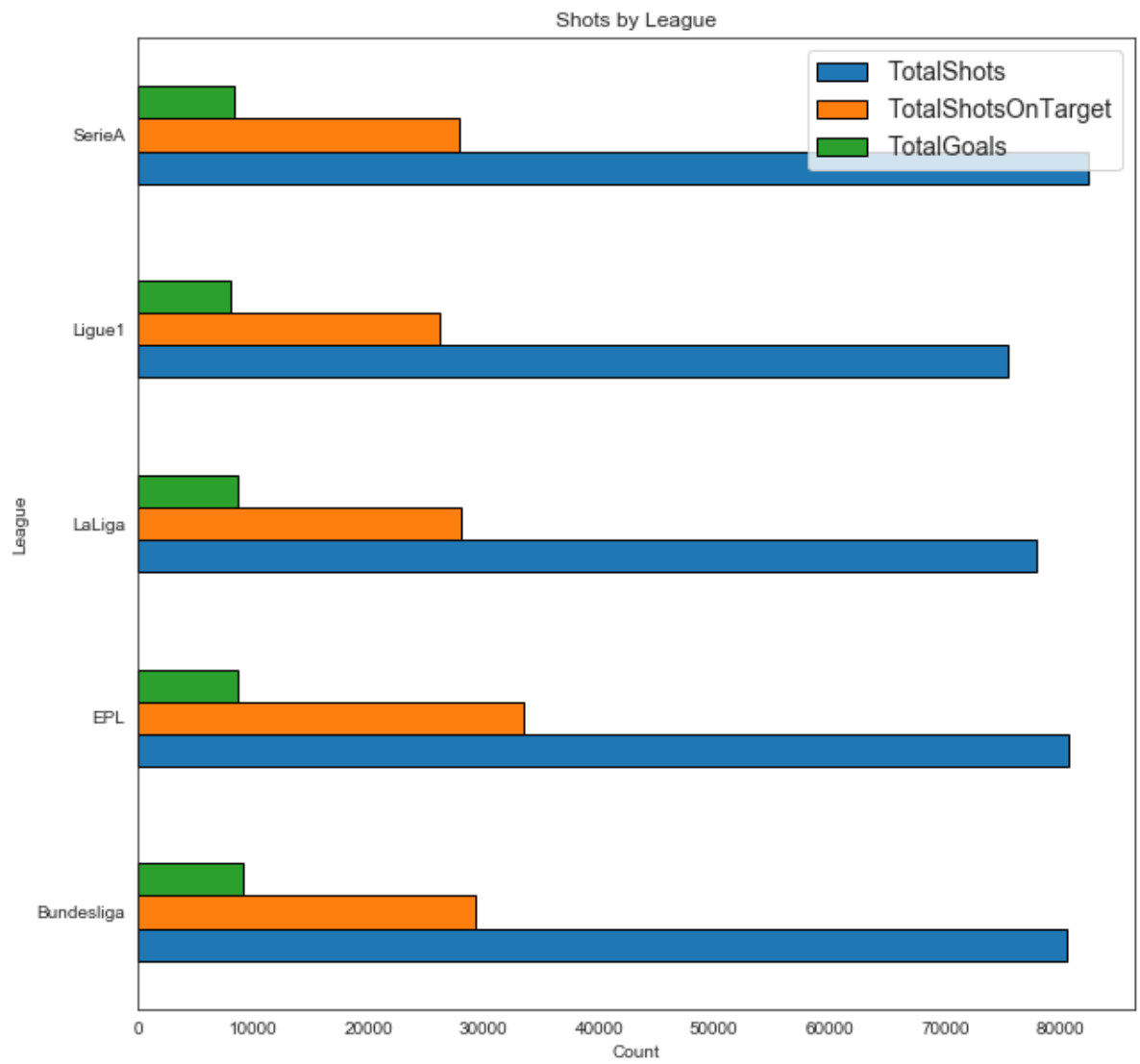
<Figure size 432x288 with 0 Axes>



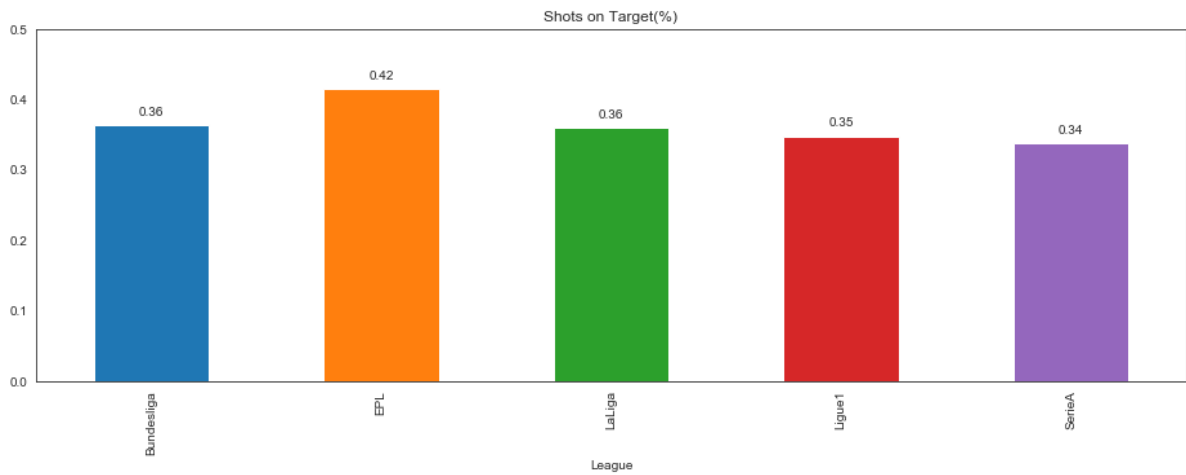
EPL has highest number of shots on target for both home and away team

Lets have a look at the Total number of shots on target as well. Total shots on target = Home team shots on target + Away team shots on target

<Figure size 432x288 with 0 Axes>

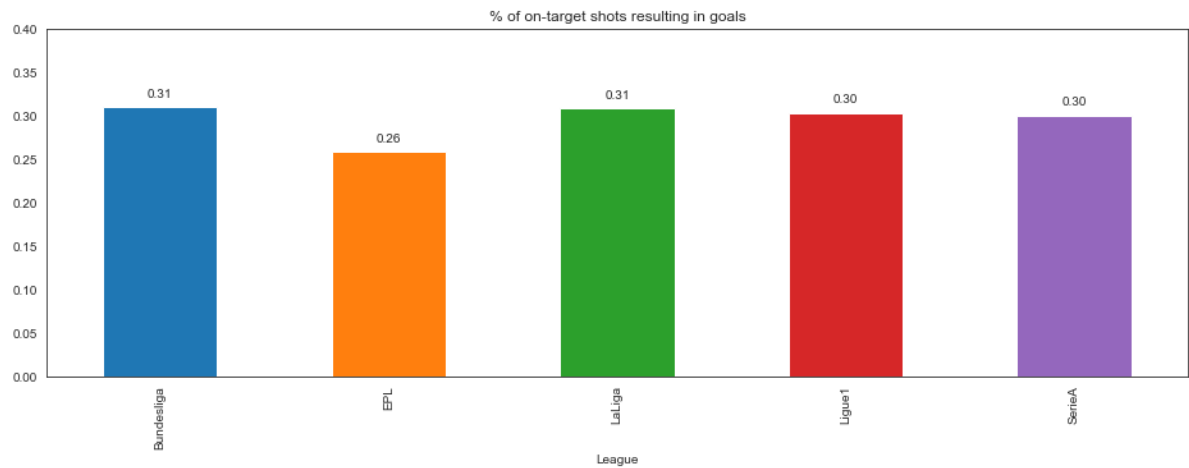


<Figure size 432x288 with 0 Axes>



EPL forwards land the maximum proportion of shots on target

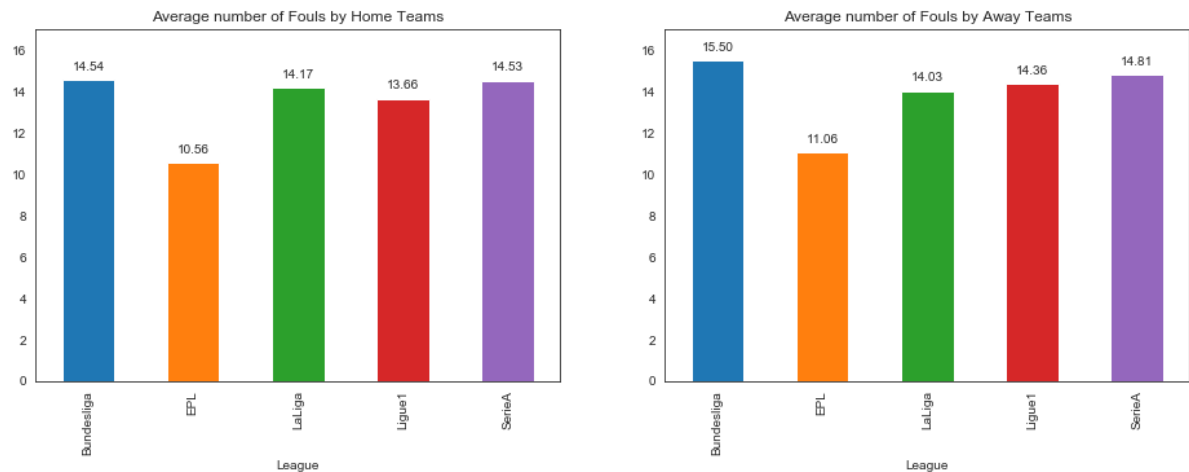
<Figure size 432x288 with 0 Axes>



However, EPL ranks behind all other leagues in converting on target shots into goals

i. Home/Away team fouls committed : Number of fouls committed by Home or Away team in a game

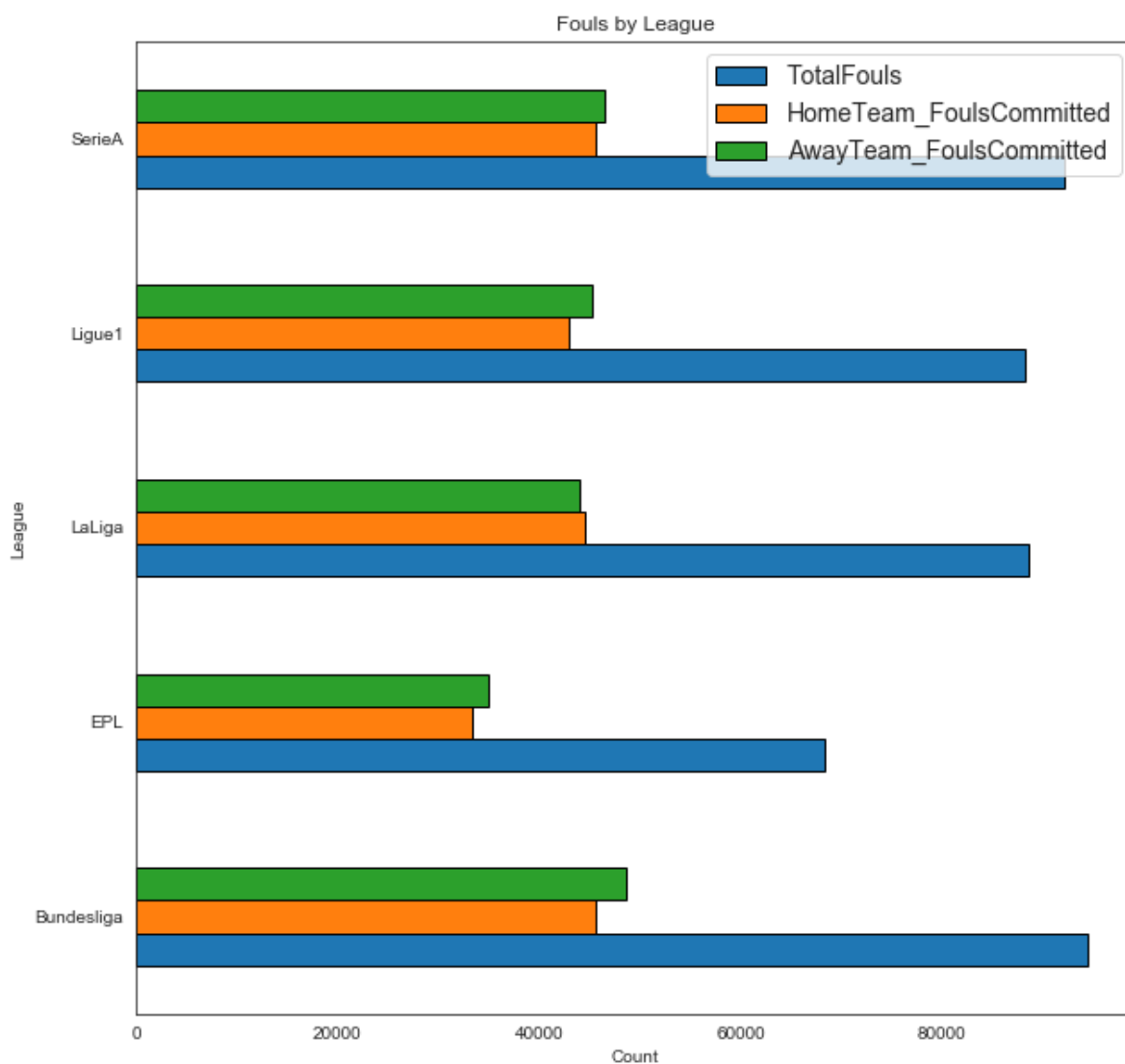
<Figure size 432x288 with 0 Axes>



EPL clearly is the most disciplined league with least number of fouls by both Home and Away teams. Away Teams in Bundesliga commit the most number of fouls per match

Lets have a look at the Total number of fouls as well. Total number of fouls = sum of fouls committed by home and away team in a match

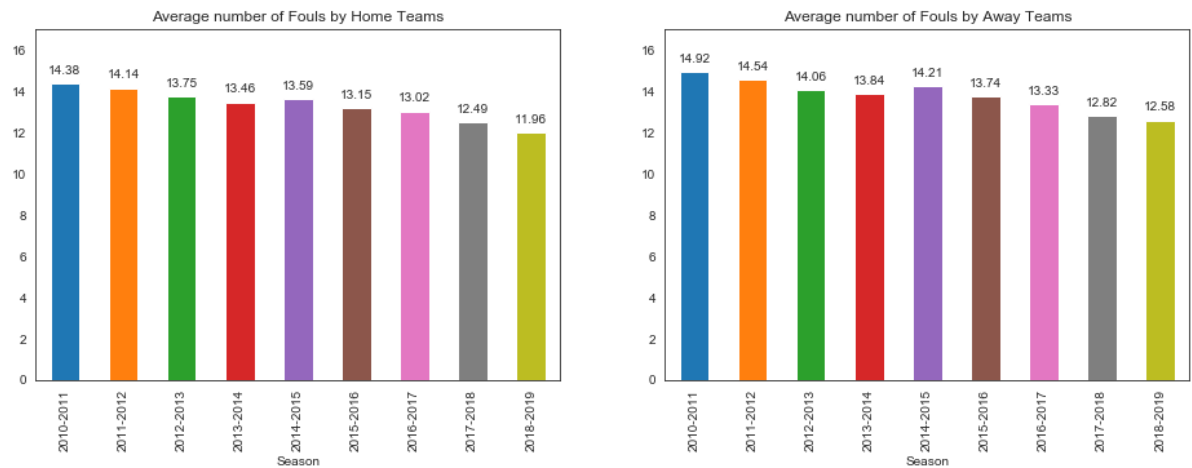
<Figure size 432x288 with 0 Axes>



Clearly EPL has least number of fouls whereas Serie A and Bundesliga(adjusted) has very high number of fouls. Premier league is well known to have "fast paced" matches, which might explain the fewer number of fouls.

Lets look the seasonal trend in number of fouls

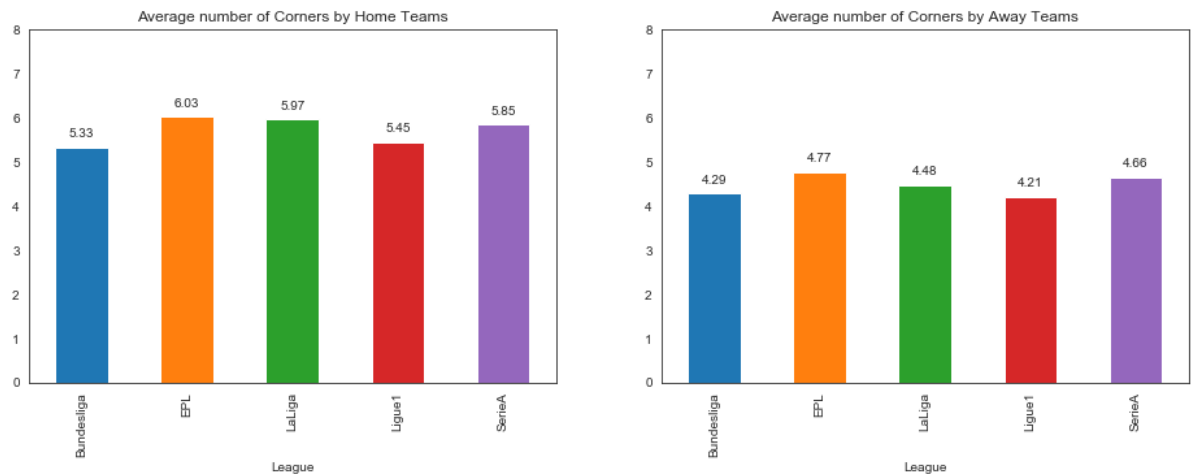
<Figure size 432x288 with 0 Axes>



The number of fouls is on a steady decline over last 9 years

j. Home/Away Team Corners : Number of corner kicks awarded to home/away teams in every game

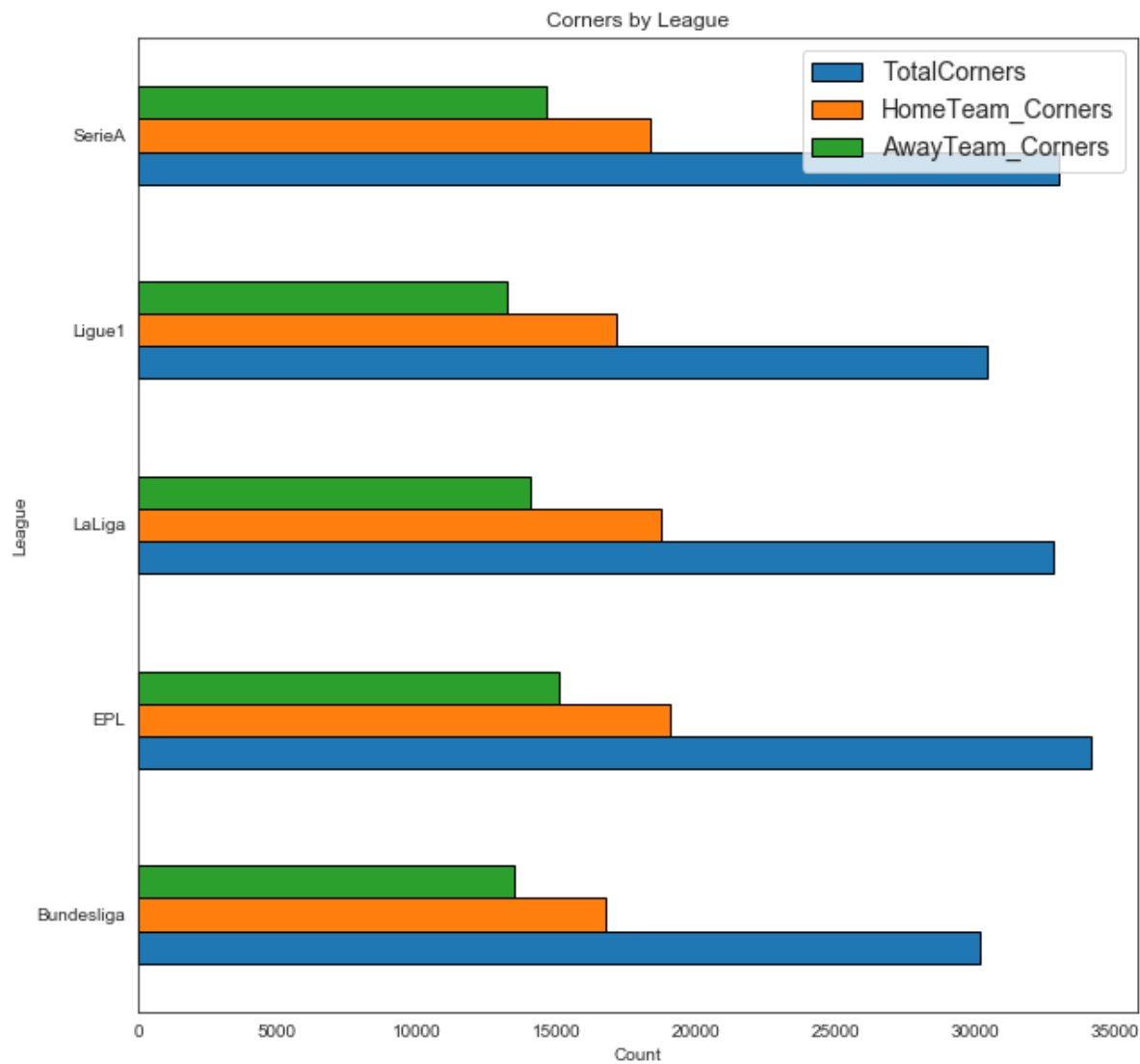
<Figure size 432x288 with 0 Axes>



Interestingly, across all leagues, home team is awarded higher number of corner kicks compared to away team (Probably indicates that home team is usually more aggressive in approach)

Lets have a look at the Total number of corner kicks as well.

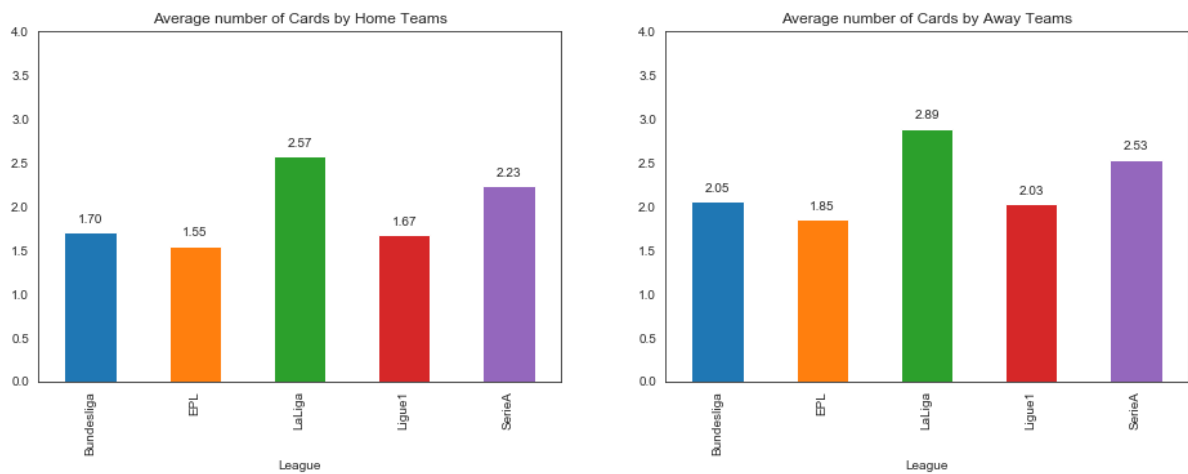
<Figure size 432x288 with 0 Axes>



EPL has very high number of corner kicks. Ligue One produces the minimum number of corner kicks

Home/Away Team Red/Yellow cards : Number of red/yellow cards shown to players from home/away team in every match

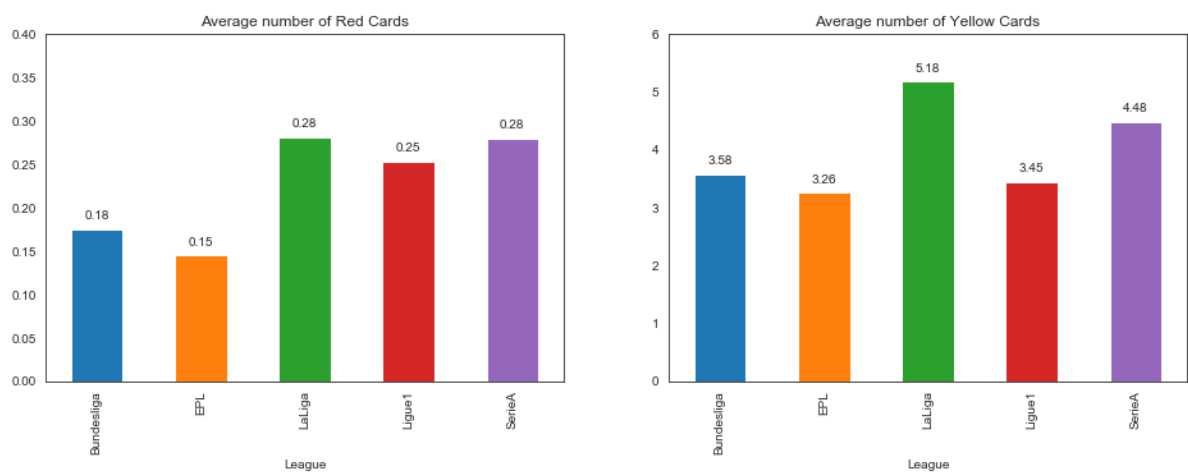
<Figure size 432x288 with 0 Axes>



Clearly, La Liga witnesses the maximum number of cards while EPL sees the minimum number of cards. Away Teams face cards more often than Home Teams

Lets take a look at variation of red and yellow card data between leagues

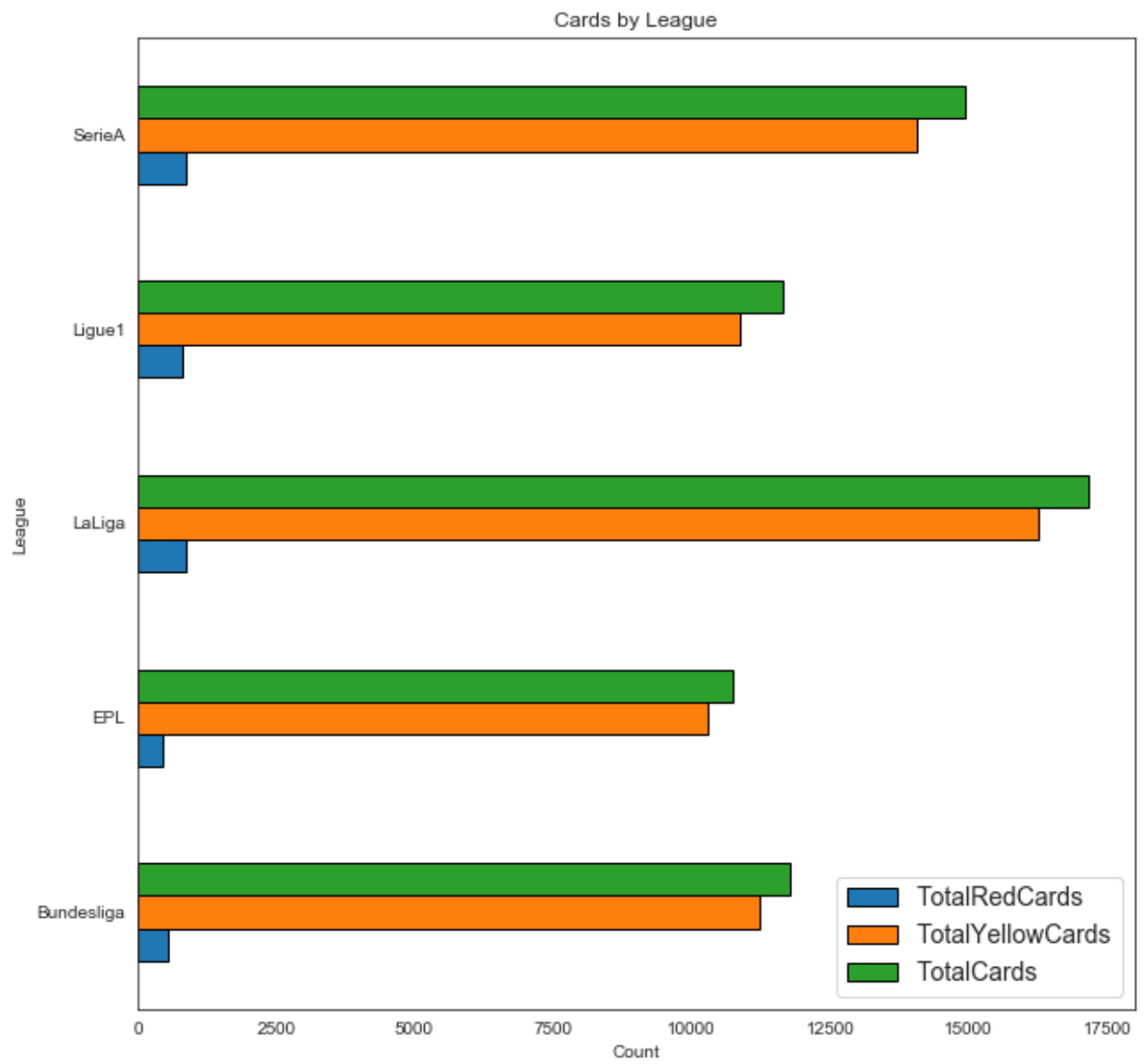
<Figure size 432x288 with 0 Axes>



Red cards are quite rare in every league (extremely rare in EPL). La Liga and Serie A has highest number of red as well as yellow cards.

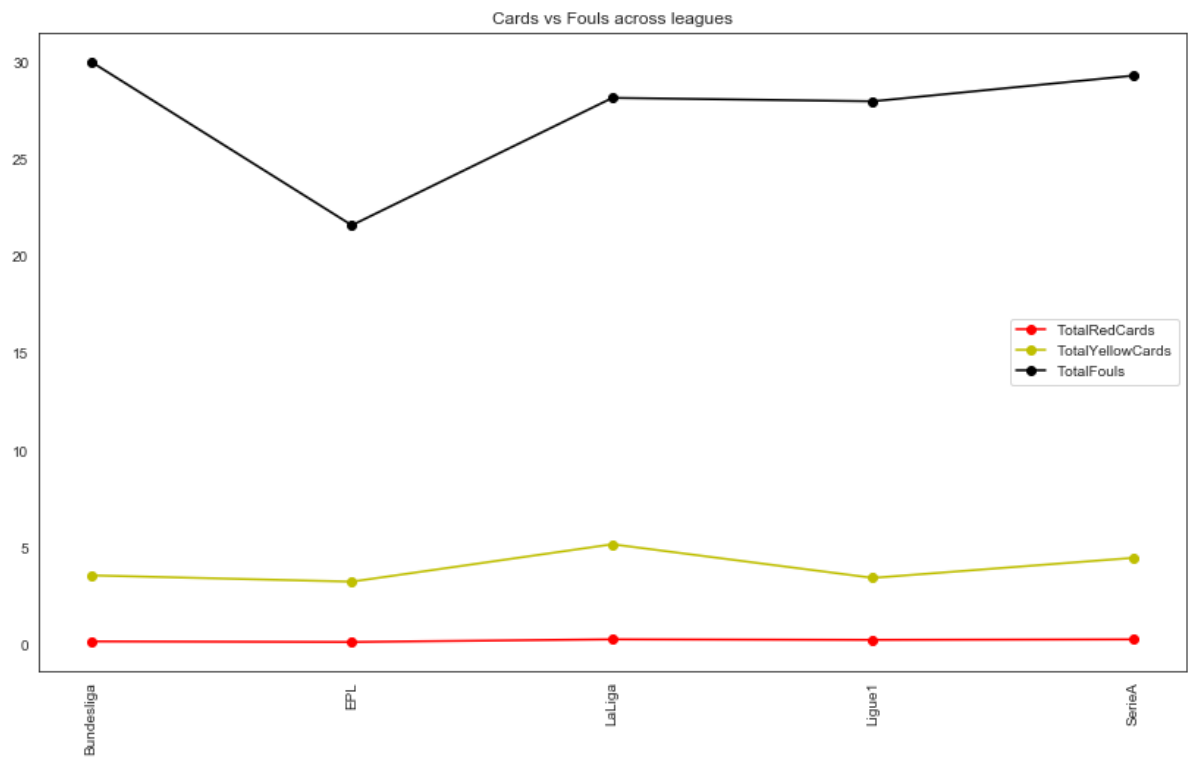
EPL witnesses least number of cards (in sync with our observation that it witnesses least number of fouls as well)

<Figure size 432x288 with 0 Axes>



Is there a correlation between number of cards and number of fouls ?

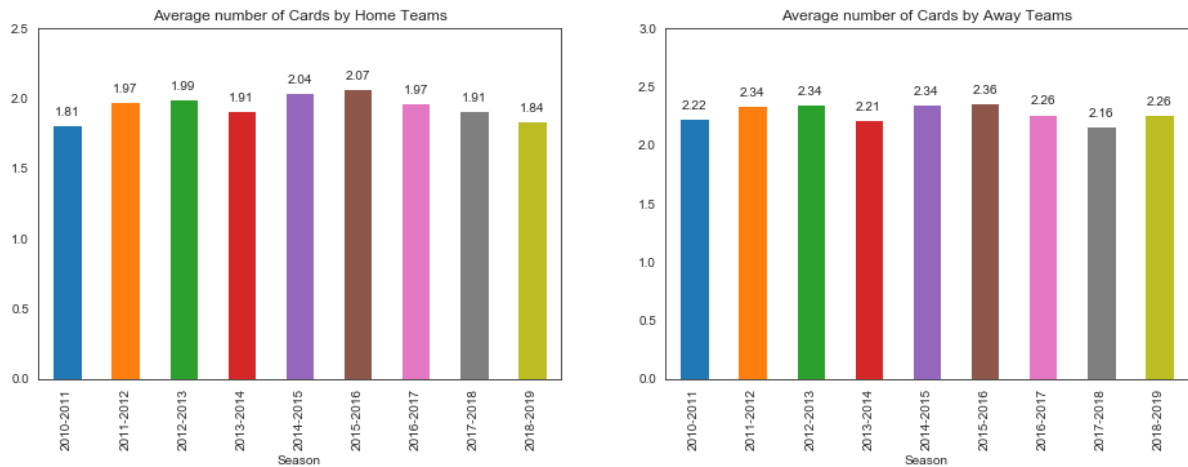
<Figure size 432x288 with 0 Axes>



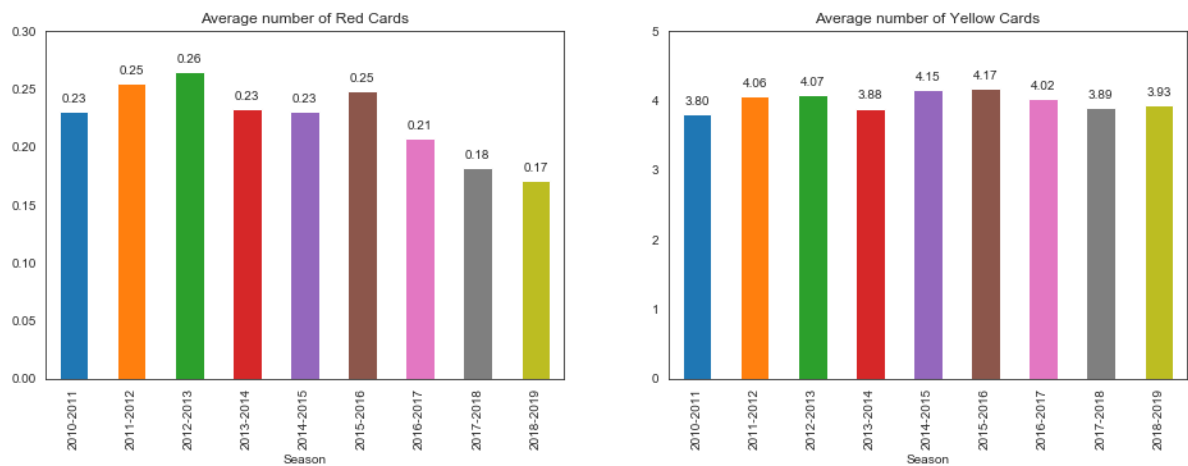
Except in EPL, there seems to be a linear correlation between number of cards and number of fouls

Lets look at how card data varied over the years

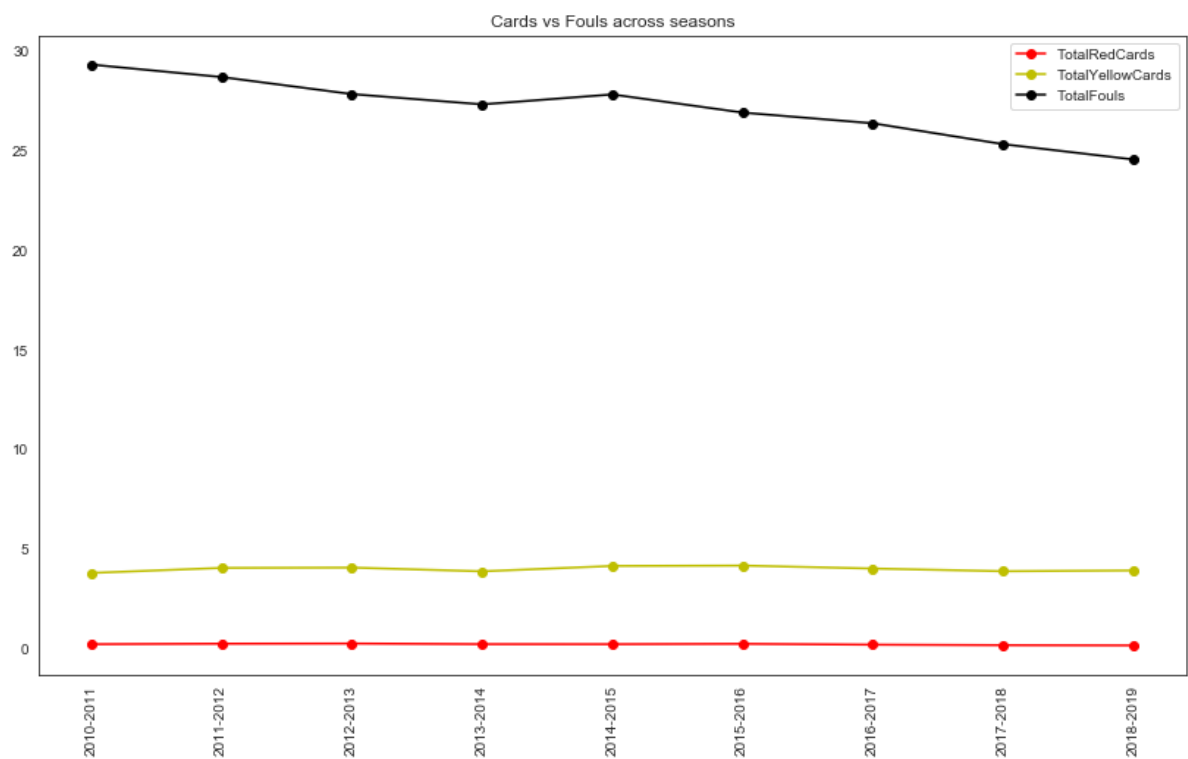
<Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>

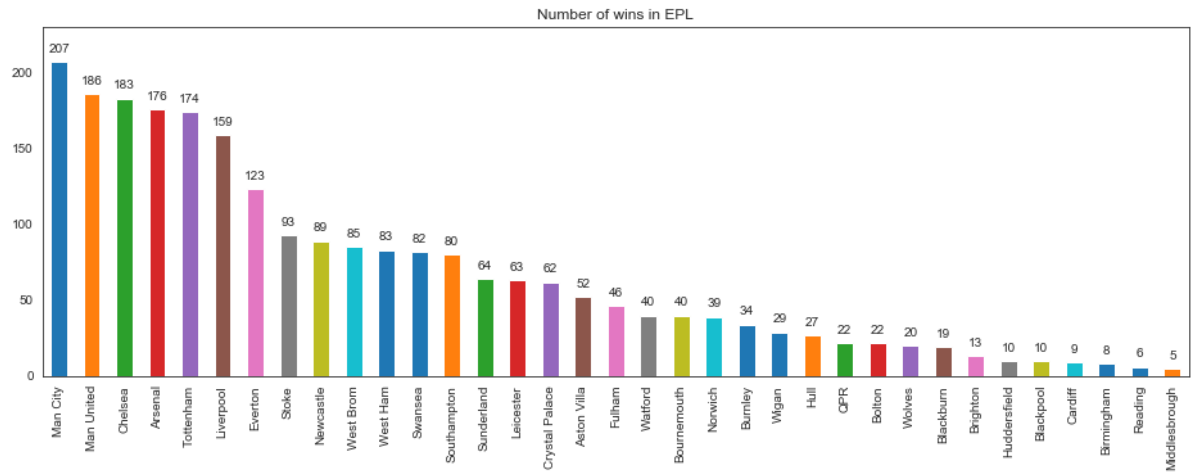


Clearly while the average number of fouls per game has come down significantly, the average number of cards havent changed much. This indicates that referees have become more strict over the years

Winner indicates the winning team

Lets look the most successful teams (teams with most number of wins in every league)

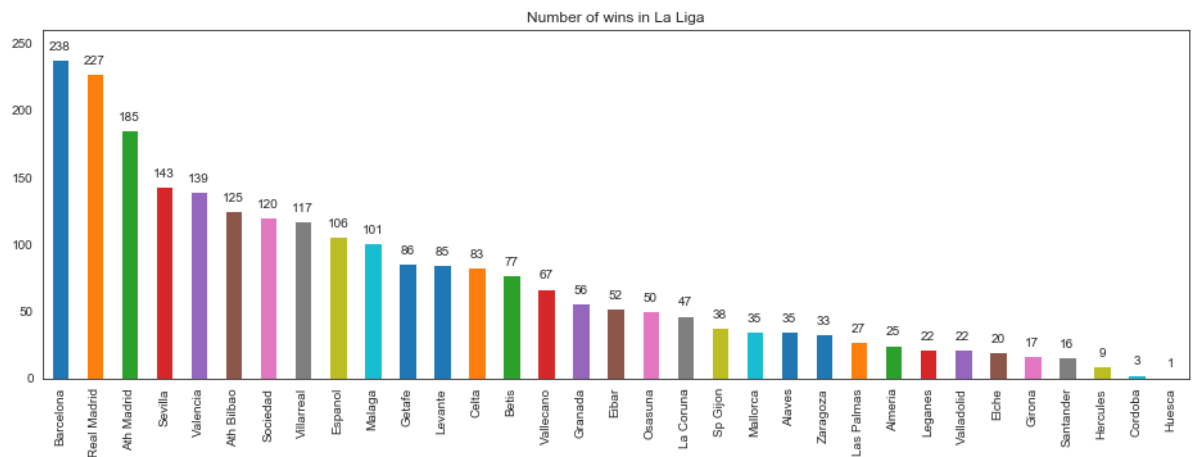
<Figure size 432x288 with 0 Axes>



Interestingly, Manchester City (and not Manchester United) has been the most successful team in EPL since 2010 The top 5 EPL teams (in terms of number of wins) are :

1. Manchester City
2. Manchester United
3. Chelsea
4. Arsenal
5. Tottenham Hotspur

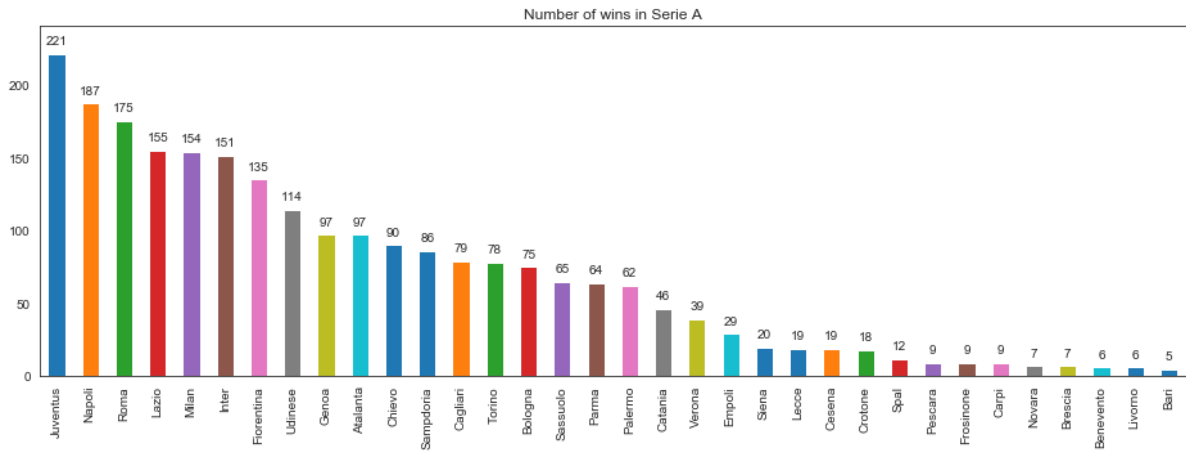
<Figure size 432x288 with 0 Axes>



FC Barcelona has been the most successful team in La Liga over last 10 years. The top5 teams in La Liga have been

1. FC Barcelona
2. Real Madrid
3. Athletic Madrid
4. Sevilla
5. Valencia

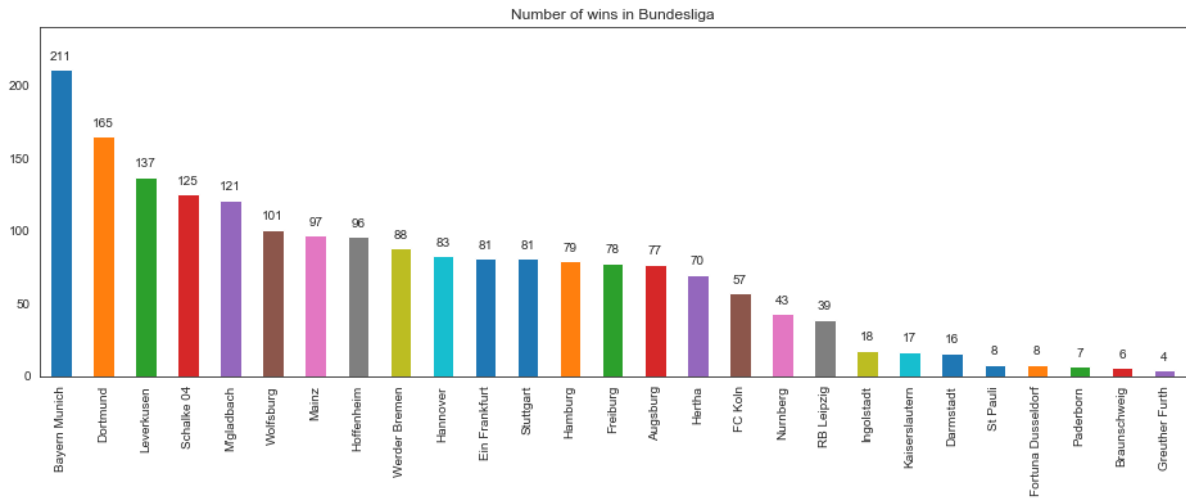
<Figure size 432x288 with 0 Axes>



Juventus has been the most successful team in Serie A over last 10 years Interesting to see AC Milan and Inter Milan at no 5 and 6 - reflecting their current decline The top 5 teams in Serie A have been :

- 1. Juventus
- 2. Napoli
- 3. Roma
- 4. Lazio
- 5. AC Milan

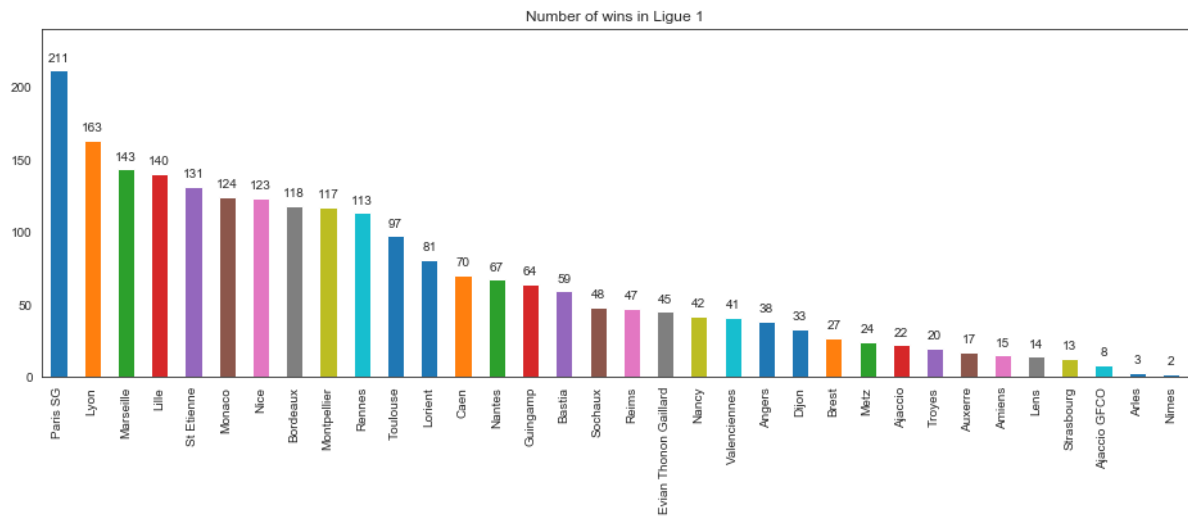
<Figure size 432x288 with 0 Axes>



As expected Bayern Munich emerges as the top team in Bundesliga over last 10 years The top 5 teams are :

- 1. Bayern Munich
- 2. Borussia Dortmund
- 3. Bayer Leverkusen
- 4. Schalke04
- 5. Borussia Mönchengladbach

<Figure size 432x288 with 0 Axes>

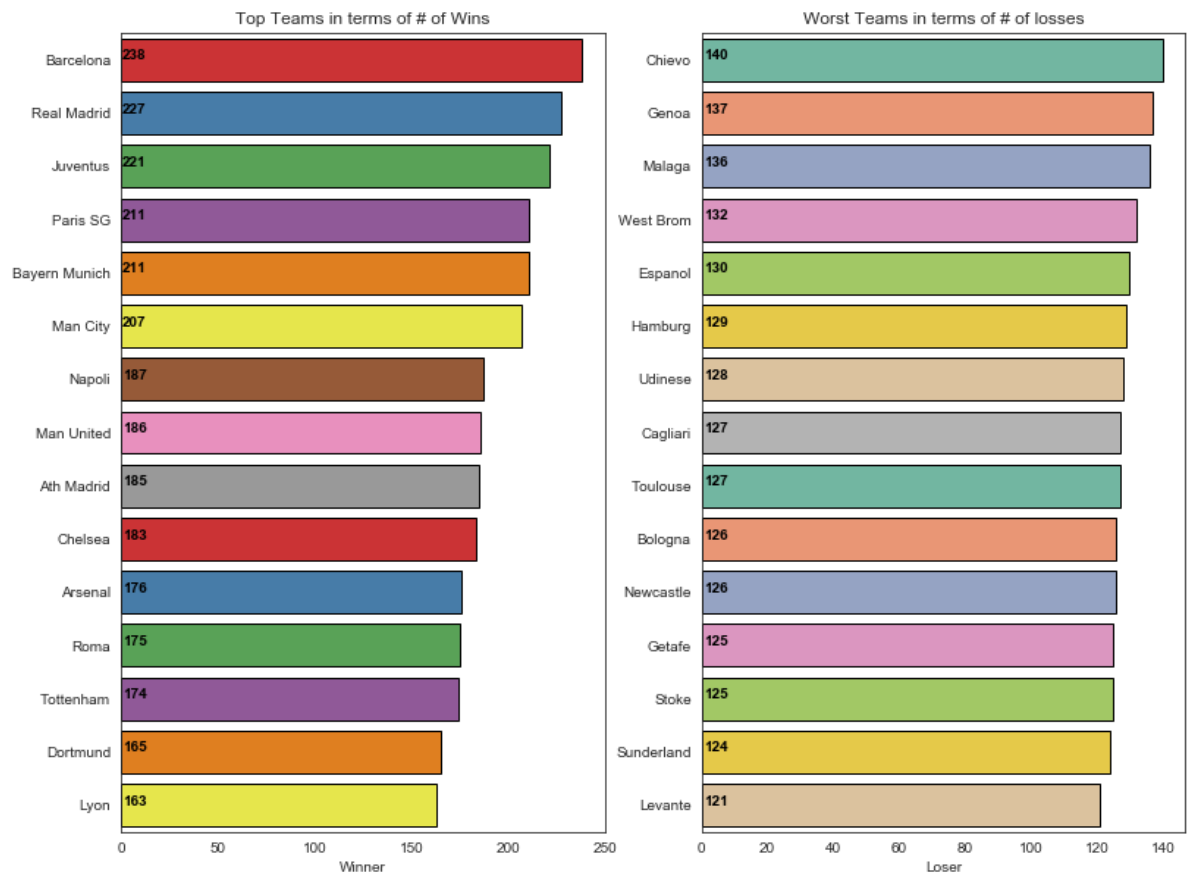


As expected, Paris Saint Germain emerges as the top team of Ligue 1 The top 5 teams in League 1 have been :

- 1. Paris Saint Germain
- 2. Lyon
- 3. Marseille
- 4. Lille
- 5. St Etienne

Overall, these are the top 15 and bottom 15 teams in terms of number of wins across all leagues

<Figure size 432x288 with 0 Axes>



So the two La Liga giants - Barcelona and Real Madrid has won the maximum number of matches in last 9 seasons

But winning isnt everything - a draw (especially an away draw) is much more valuable than a loss

Let's implement a points system :

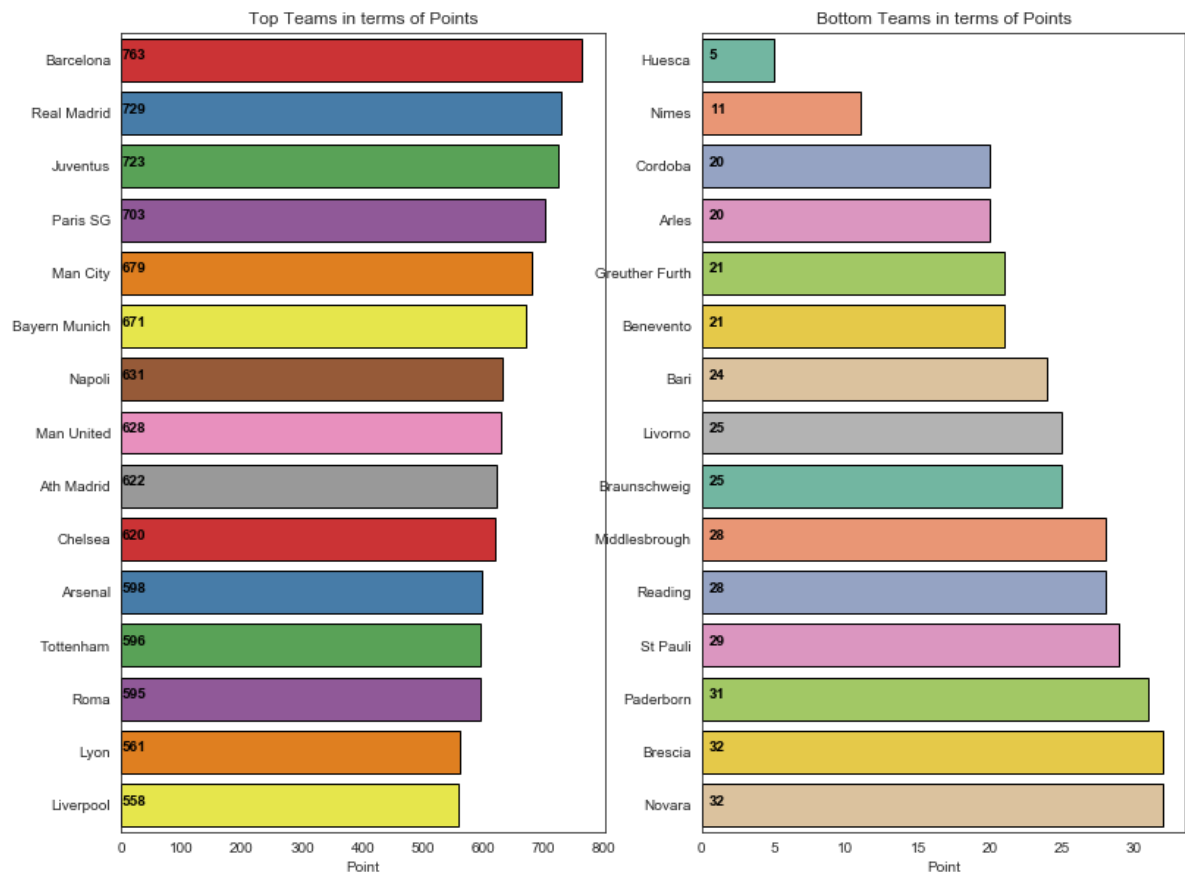
3 points for a win

1 point for a draw

0 point for a loss

Now lets have a look at top teams on the basis of points

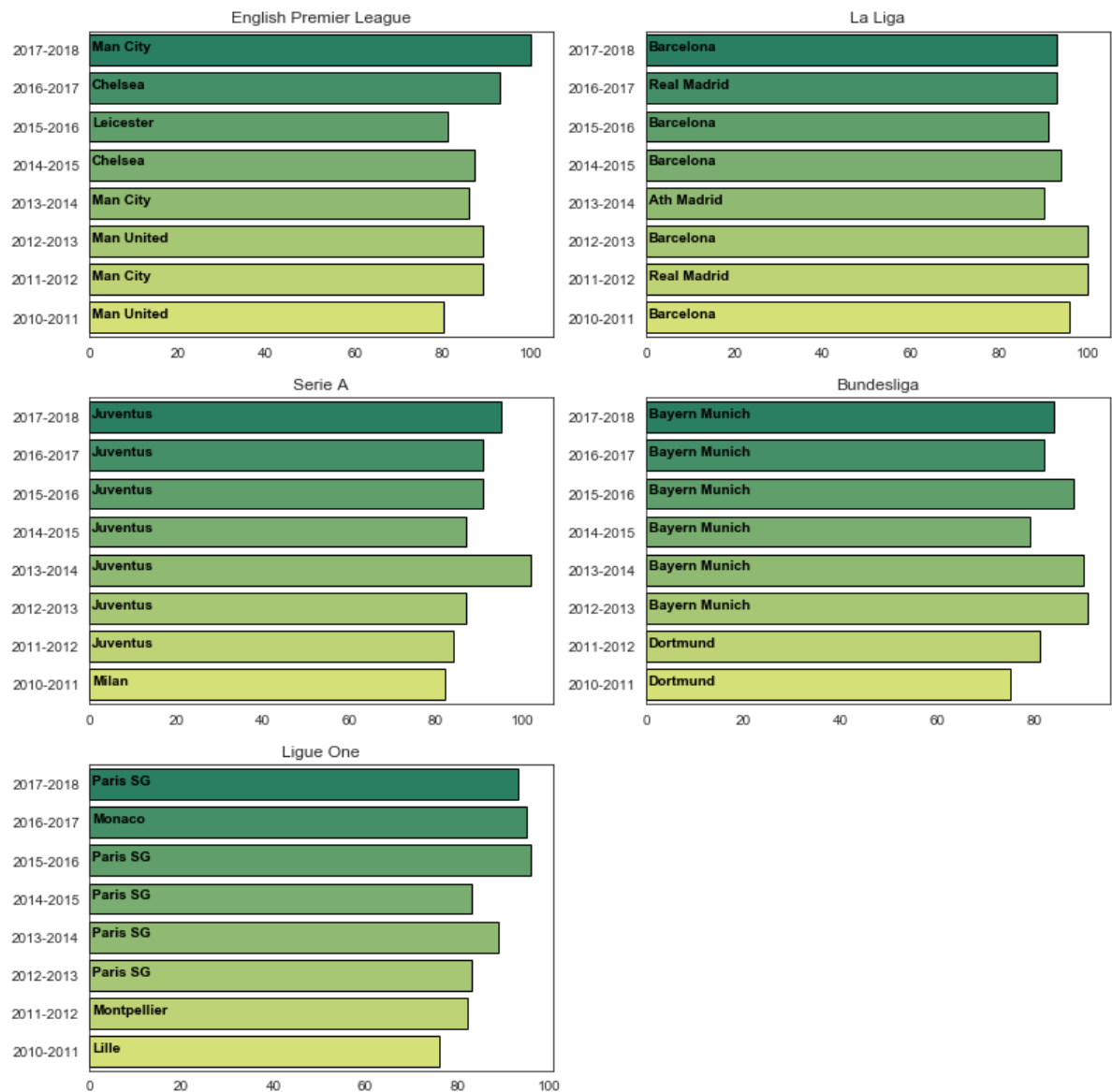
<Figure size 432x288 with 0 Axes>



We still have FC Barcelona and Real Madrid at the top

Lets also look at the top performers in every league and season based on points. We'll exclude current ongoing season as it has incomplete data

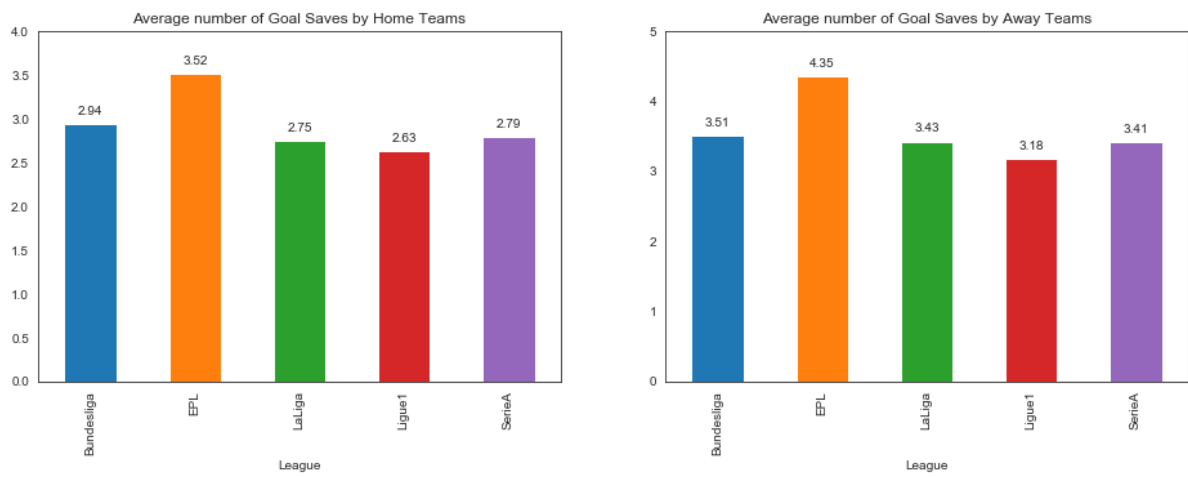
<Figure size 432x288 with 0 Axes>



In most leagues its generally one or two clubs dominating the league. The only exception is EPL where we had 4 winners in last 8 years

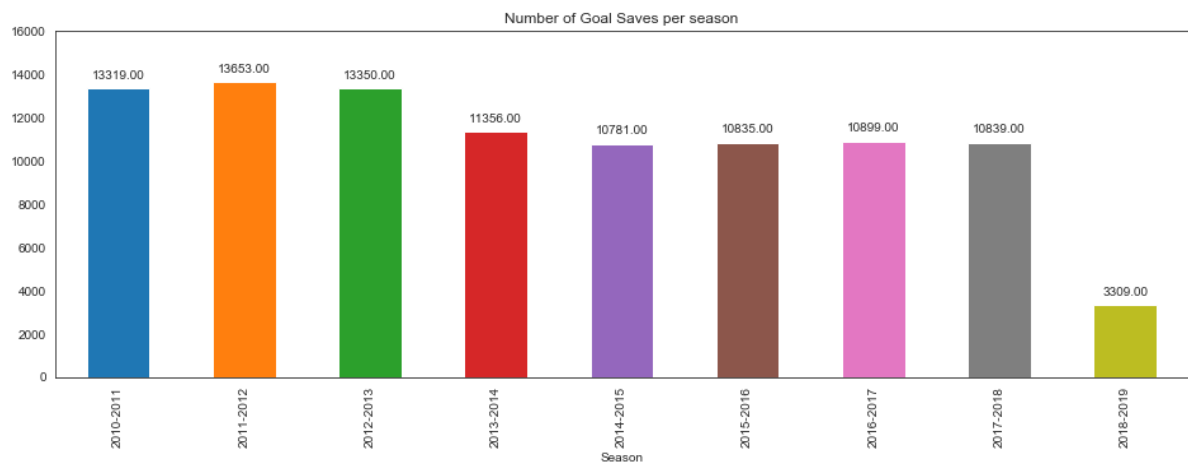
Home/Away Goal Saves : The number of shots that were on target but didn't result in a goal (Not necessarily all of them are saved by GoalKeeper, some might be disqualified due to offside rule)

<Figure size 432x288 with 0 Axes>



Interestingly Away teams save more goals than home team in almost every league. But that might also be because home teams usually are more effective with number of shots on target. EPL probably keeps goalkeepers busier than other leagues

<Figure size 432x288 with 0 Axes>



So from feature analysis we can conclude the following :

1. EPL is the most volatile league in terms of league qualification. Only 35 % of teams have been consistently present in last 10 years
2. There's a significant home advantage in every league. Away performance is improving in general but not at a very rapid pace
3. EPL and LaLiga produces more goals compared to other leagues
4. Both Home and Away teams produce more goals during second half
5. Forwards in La Liga and Bundesliga are the most effective. Forwards in Serie A are the least effective
6. EPL clearly is the most disciplined and least violent league with least number of fouls and cards. La Liga leads in number of red and yellow cards
7. Average number of fouls have fallen across all leagues over last 10 years. However average number of cards haven't
8. EPL forwards land maximum number of shots on target. However, EPL goalkeepers save maximum number of goals as well

Features and their impact on match outcome

For correlation analysis, lets convert the match results into numeric values

We also want to check if higher number of attempts have a correlation with wins (Is it important to be more aggressive ? or does it pay to be more accurate ?)

Lets create an extra feature called Winner_Shots. This will contain the number of attempts by the winning team

Similarly we can create a feature called Winner Accuracy. This will contain the proportion of shots by winning team that are on target

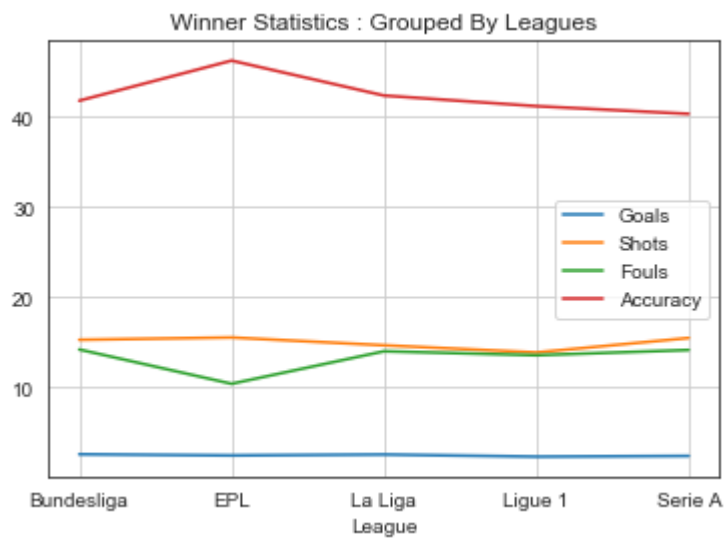
Similarly we can create a feature called Winner Goals

And also a feature called Winner Fouls

Lets have a look at the winner statistics, grouped by league

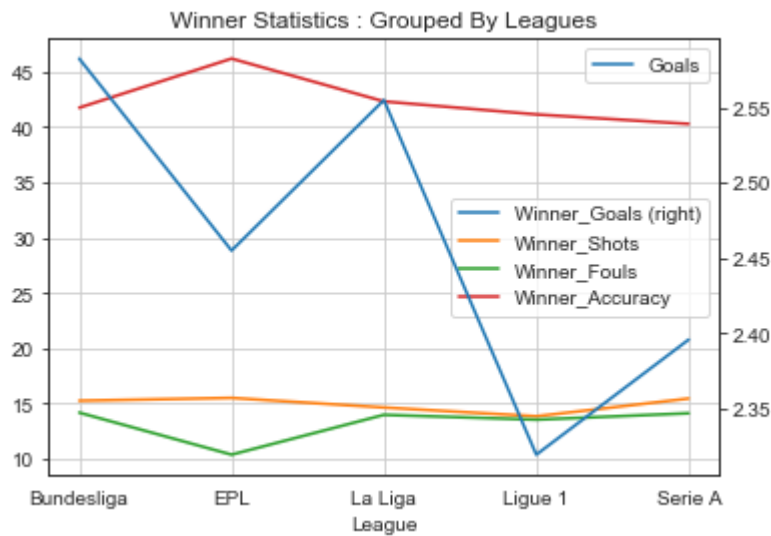
| | Winner_Goals | Winner_Shots | Winner_Fouls | Winner_Accuracy |
|-------------------|--------------|--------------|--------------|-----------------|
| League | | | | |
| Bundesliga | 2.582399 | 15.264884 | 14.191544 | 41.749881 |
| EPL | 2.454735 | 15.518106 | 10.390669 | 46.191782 |
| LaLiga | 2.554817 | 14.653821 | 13.990698 | 42.320281 |
| Ligue1 | 2.319014 | 13.857746 | 13.553345 | 41.146722 |
| SerieA | 2.395528 | 15.464011 | 14.132075 | 40.289794 |

<Figure size 432x288 with 0 Axes>



Because the goal data has lesser variation, its not very visible in this graph. Lets use a different scale just for the goal data

<Figure size 432x288 with 0 Axes>



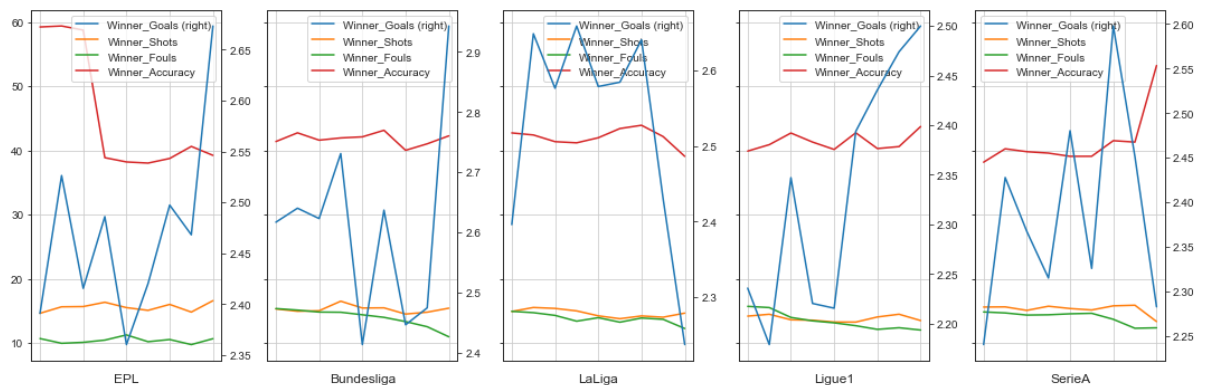
The number goals per game and shots per game of winner are steady which is 2.27 - 2.52 and 13.1 - 14.60 respectively. Bundesliga has the highest goal per game for winners. The EPL has the lowest number of fouls committed for winner. This may caused by higher tolerance of fouls in the EPL.

Now we would like to study the change for those 3 parameters for winners across different seasons.

Lets group each league data by seasons and look at the visual representation

<Figure size 432x288 with 0 Axes>

C:\Users\ksaha\AppData\Local\Continuum\anaconda3\lib\site-packages\matplotlib\figure.py:2299: UserWarning: This figure includes Axes that are not compatible with tight_layout, so results might be incorrect.
 warnings.warn("This figure includes Axes that are not compatible "



Again, goal is steady for each league and across all seasons.

However, the number of shots fluctuated a lot for every league. La liga has significant drop in the number of shot per game.

The most interesting feature is the fouls committed per game. Except EPL, all other 4 leagues has significant drop in the number of fouls committed per game over seasons.

Next, we would like compare some of key features that can affect the game result by answering following questions:

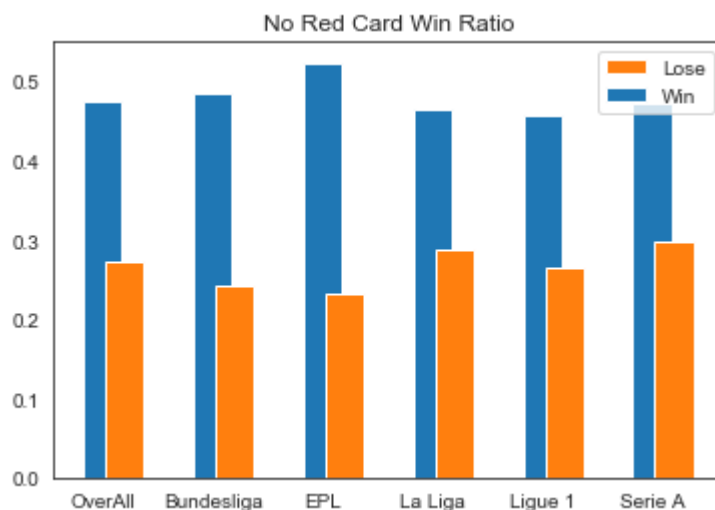
- 1.What's the effect of red card on the win ratio?
- 2.What's the effect of number of shots on the win ratio?
- 3.What's the effect of number of fouls committed on the win ratio?
- 4.What's the effect of corners on the win ratio?

Let's start with the effect of red card

Calculate the win ratio of with and without red card group by league.

Out[160]:

| | RedResult | noRedResult | RedCardWinRatio | NoRedCardWinRatio |
|-------------------|-----------|-------------|-----------------|-------------------|
| League | | | | |
| Bundesliga | 104 | 208 | 0.242991 | 0.485981 |
| EPL | 103 | 232 | 0.232506 | 0.523702 |
| LaLiga | 236 | 382 | 0.287454 | 0.465286 |
| Ligue1 | 197 | 339 | 0.265499 | 0.456873 |
| SerieA | 244 | 387 | 0.297561 | 0.471951 |

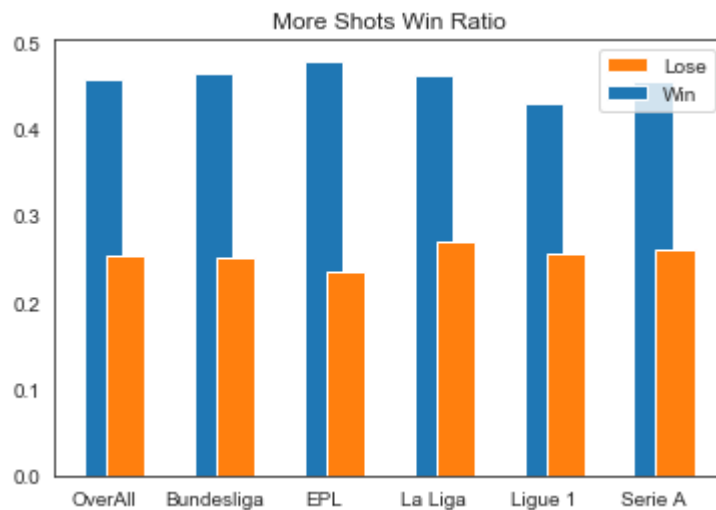


We can clearly see that the red card has huge impact on the win ratio, almost doubled the win ratio, especially in the EPL.

Then let's repeat the same procedure with the number of shots per game.

Out[165]:

| | moreshots | lessshots | MoreshotsWinRatio | LessshotsWinRatio |
|-------------------|-----------|-----------|-------------------|-------------------|
| League | | | | |
| Bundesliga | 1176 | 638 | 0.465006 | 0.252274 |
| EPL | 1512 | 742 | 0.478481 | 0.234810 |
| LaLiga | 1448 | 847 | 0.461146 | 0.269745 |
| Ligue1 | 1354 | 804 | 0.429841 | 0.255238 |
| SerieA | 1430 | 818 | 0.455414 | 0.260510 |

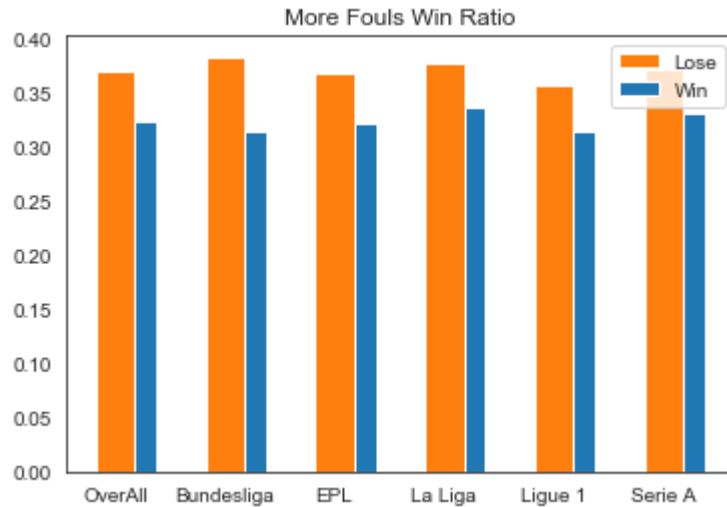


As we expect, the team with more shots tends to win the game. However, the effect is smaller than the effect of red card.

Repeat the procedure with numbers of fouls committed.

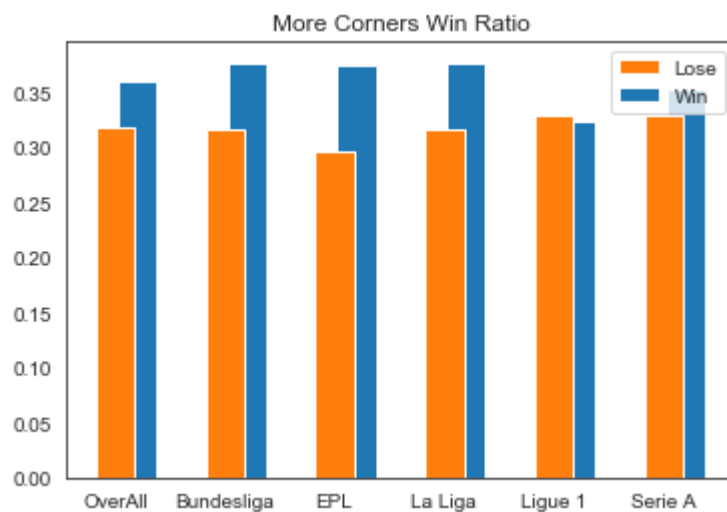
Out[170]:

| | moreFouls | lessFouls | MoreFoulsSTD | MoreFoulsWinRatio | LessFoulsWinRatio |
|-------------------|-----------|-----------|--------------|-------------------|-------------------|
| League | | | | | |
| Bundesliga | 795 | 969 | 0.464350 | 0.314353 | 0.383155 |
| EPL | 1015 | 1160 | 0.467012 | 0.321203 | 0.367089 |
| LaLiga | 1059 | 1183 | 0.472850 | 0.337261 | 0.376752 |
| Ligue1 | 991 | 1122 | 0.464431 | 0.314603 | 0.356190 |
| SerieA | 1037 | 1168 | 0.470380 | 0.330255 | 0.371975 |



The effect of fouls has small effect on the win ratio. However, the team committing more fouls tends to lose the game, This could be because of the fact that these teams have high pressure on defense.

Repeat the procedure with corners.



In general, more corners tend to result in wins. But this is not the case for Ligue 1.

What are the things that league winning teams tend to do differently ? Let's try to find an answer to this

We'll use the point metric to find the winning team

Out[178]:

| | League | HomeTeam | Season | HomeTeam_Points | AwayTeam_Points |
|---|--------|-------------|-----------|-----------------|-----------------|
| 0 | LaLiga | Almeria | 2010-2011 | 19 | 11 |
| 1 | Ligue1 | Arles | 2010-2011 | 12 | 8 |
| 2 | EPL | Arsenal | 2010-2011 | 37 | 31 |
| 3 | EPL | Aston Villa | 2010-2011 | 31 | 17 |
| 4 | LaLiga | Ath Bilbao | 2010-2011 | 37 | 21 |

Lets find total points for every team in every season

Out[181]:

| | | | TotalPoints |
|------------|-----------|---------------|-------------|
| League | Season | HomeTeam | |
| Bundesliga | 2010-2011 | Bayern Munich | 65 |
| | | Dortmund | 75 |
| | | Ein Frankfurt | 34 |
| | | FC Koln | 44 |
| | | Freiburg | 44 |

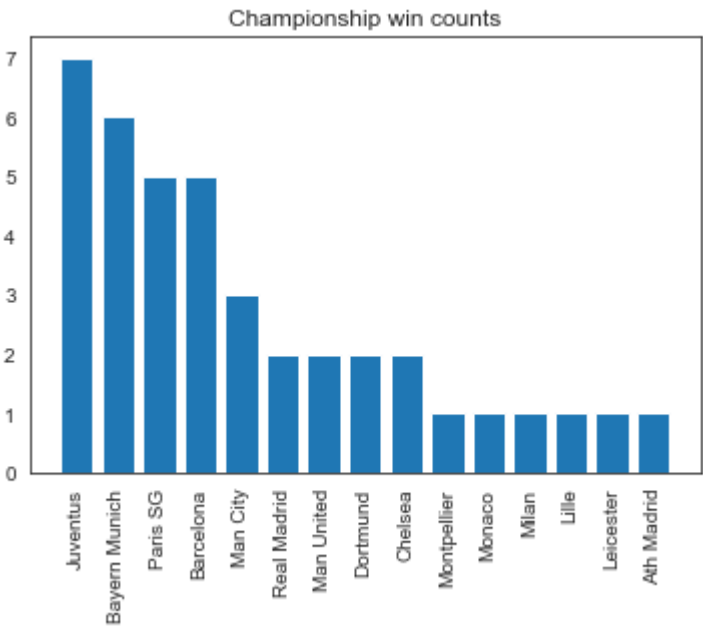
Find the champion for each league for each season.

Out[182]:

| | League | Season | Team |
|----|------------|-----------|---------------|
| 0 | Bundesliga | 2010-2011 | Dortmund |
| 1 | EPL | 2010-2011 | Man United |
| 2 | LaLiga | 2010-2011 | Barcelona |
| 3 | Ligue1 | 2010-2011 | Lille |
| 4 | SerieA | 2010-2011 | Milan |
| 5 | Bundesliga | 2011-2012 | Dortmund |
| 6 | EPL | 2011-2012 | Man City |
| 7 | LaLiga | 2011-2012 | Real Madrid |
| 8 | Ligue1 | 2011-2012 | Montpellier |
| 9 | SerieA | 2011-2012 | Juventus |
| 10 | Bundesliga | 2012-2013 | Bayern Munich |
| 11 | EPL | 2012-2013 | Man United |
| 12 | LaLiga | 2012-2013 | Barcelona |
| 13 | Ligue1 | 2012-2013 | Paris SG |
| 14 | SerieA | 2012-2013 | Juventus |
| 15 | Bundesliga | 2013-2014 | Bayern Munich |
| 16 | EPL | 2013-2014 | Man City |
| 17 | LaLiga | 2013-2014 | Ath Madrid |
| 18 | Ligue1 | 2013-2014 | Paris SG |
| 19 | SerieA | 2013-2014 | Juventus |
| 20 | Bundesliga | 2014-2015 | Bayern Munich |
| 21 | EPL | 2014-2015 | Chelsea |
| 22 | LaLiga | 2014-2015 | Barcelona |
| 23 | Ligue1 | 2014-2015 | Paris SG |
| 24 | SerieA | 2014-2015 | Juventus |
| 25 | Bundesliga | 2015-2016 | Bayern Munich |
| 26 | EPL | 2015-2016 | Leicester |
| 27 | LaLiga | 2015-2016 | Barcelona |
| 28 | Ligue1 | 2015-2016 | Paris SG |
| 29 | SerieA | 2015-2016 | Juventus |
| 30 | Bundesliga | 2016-2017 | Bayern Munich |
| 31 | EPL | 2016-2017 | Chelsea |
| 32 | LaLiga | 2016-2017 | Real Madrid |
| 33 | Ligue1 | 2016-2017 | Monaco |
| 34 | SerieA | 2016-2017 | Juventus |

| | League | Season | Team |
|----|------------|-----------|---------------|
| 35 | Bundesliga | 2017-2018 | Bayern Munich |
| 36 | EPL | 2017-2018 | Man City |
| 37 | LaLiga | 2017-2018 | Barcelona |
| 38 | Ligue1 | 2017-2018 | Paris SG |
| 39 | SerieA | 2017-2018 | Juventus |
| 40 | Bundesliga | 2018-2019 | Dortmund |
| 41 | EPL | 2018-2019 | Man City |
| 42 | LaLiga | 2018-2019 | Barcelona |
| 43 | Ligue1 | 2018-2019 | Paris SG |
| 44 | SerieA | 2018-2019 | Juventus |

```
Out[184]: Team
Juventus      7
Bayern Munich 6
Paris SG      5
Barcelona     5
Man City      3
Real Madrid   2
Man United    2
Dortmund      2
Chelsea       2
Montpellier   1
Monaco        1
Milan         1
Lille         1
Leicester     1
Ath Madrid    1
Name: Season, dtype: int64
```



Find the key features of champion team and merge into one dataframe

Out[187]:

| | | | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | HalfTi |
|-----------|------------|----------------|-------------------------|-------------------------|--------|
| Season | League | HomeTeam | | | |
| 2010-2011 | Bundesliga | Bayern Munich | 48 | 13 | |
| | | Dortmund | 35 | 8 | |
| | | Ein Frankfurt | 13 | 24 | |
| | | FC Koln | 30 | 21 | |
| | | Freiburg | 24 | 24 | |
| | | Hamburg | 29 | 24 | |
| | | Hannover | 32 | 17 | |
| | | Hoffenheim | 28 | 21 | |
| | | Kaiserslautern | 25 | 19 | |
| | | Leverkusen | 33 | 24 | |

Drop current season due to incomplete data.

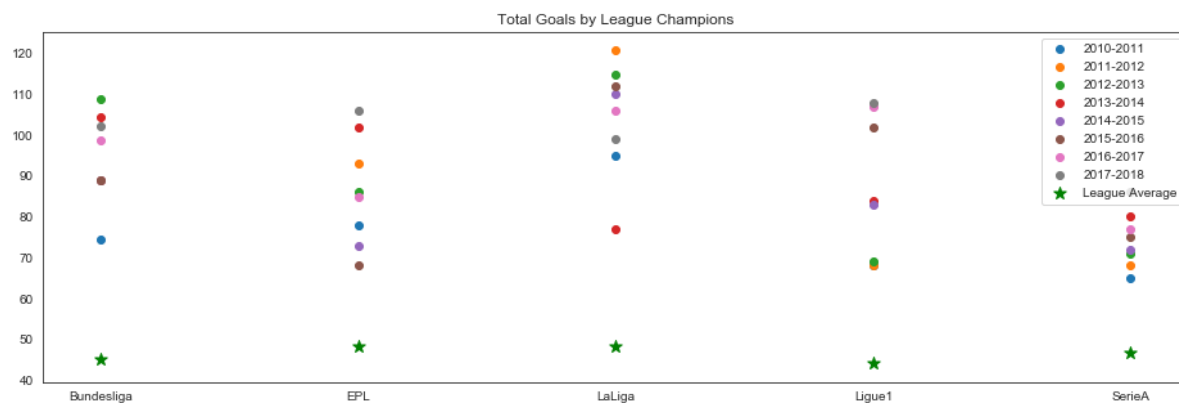
Out[195]:

| | HomeTeam | League | Season | FullTime_HomeTeam_Goals | FullTime_AwayTeam_Goals | HalfTi |
|---|------------|------------|-----------|-------------------------|-------------------------|--------|
| 0 | Dortmund | Bundesliga | 2010-2011 | 38.888889 | 35.555556 | |
| 1 | Man United | EPL | 2010-2011 | 49.000000 | 29.000000 | |
| 2 | Barcelona | LaLiga | 2010-2011 | 46.000000 | 49.000000 | |
| 3 | Lille | Ligue1 | 2010-2011 | 40.000000 | 28.000000 | |
| 4 | Milan | SerieA | 2010-2011 | 42.000000 | 23.000000 | |

Plot the total number of goals for each champion in different league.

```
Out[197]: ['2010-2011',
            '2011-2012',
            '2012-2013',
            '2013-2014',
            '2014-2015',
            '2015-2016',
            '2016-2017',
            '2017-2018',
            'League Average']
```

<Figure size 432x288 with 0 Axes>



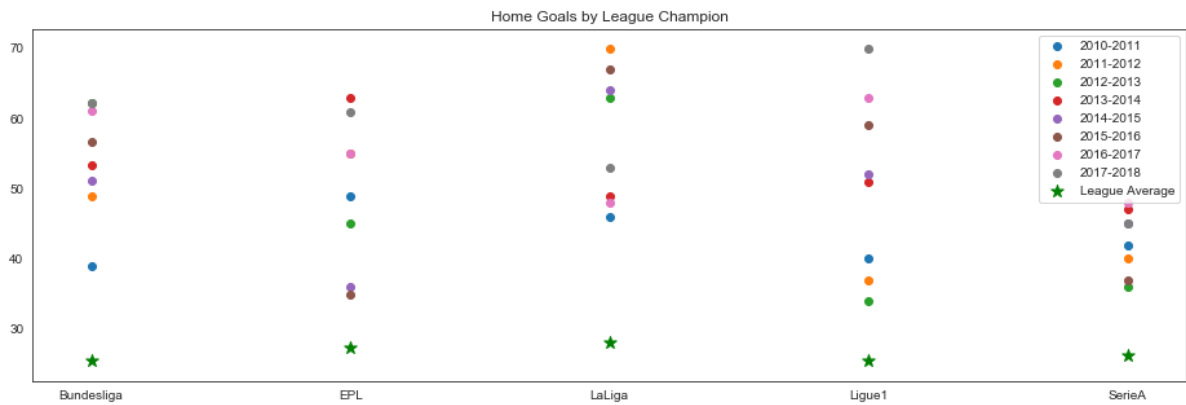
From the figure above, we can see that Laliga has the highest number of goals for champion except for one season. Serie A has the largest distribution on the total goals. The Bundesliga has the lowest number of goals since they has less games to play.

```
Out[199]: League
Bundesliga    94.444444
EPL           86.375000
LaLiga       104.375000
Ligue1       86.125000
SerieA       74.250000
dtype: float64
```

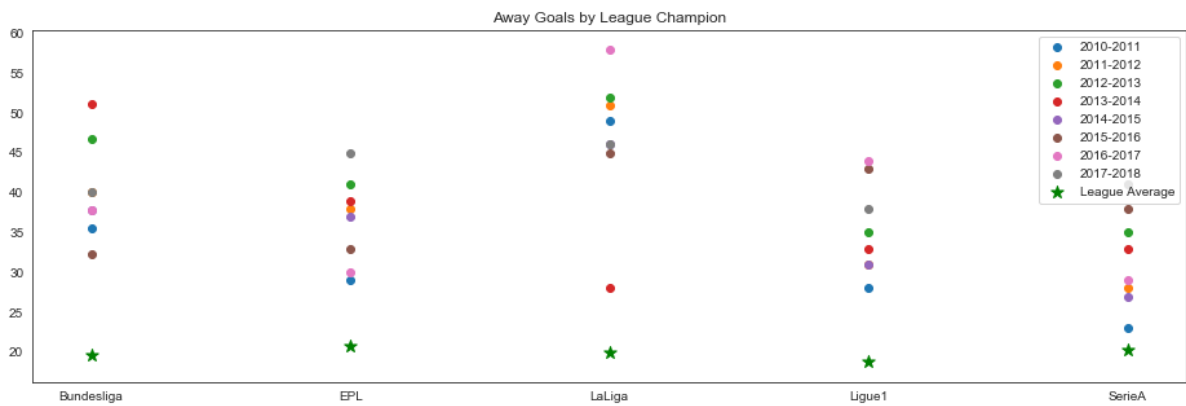
```
Out[200]: League
Bundesliga    45.228395
EPL           48.094444
LaLiga       48.161111
Ligue1       44.294444
SerieA       46.538889
dtype: float64
```

The total goal of champion team is way above the average goal of all teams.

<Figure size 432x288 with 0 Axes>

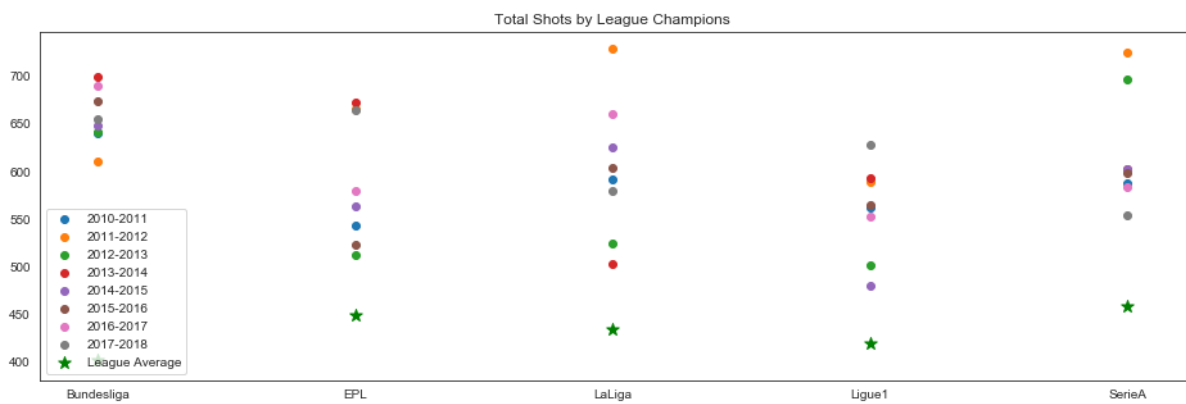


<Figure size 432x288 with 0 Axes>



Break down the goal by goals at home and away. The general trend is similar to the overall goal.

<Figure size 432x288 with 0 Axes>



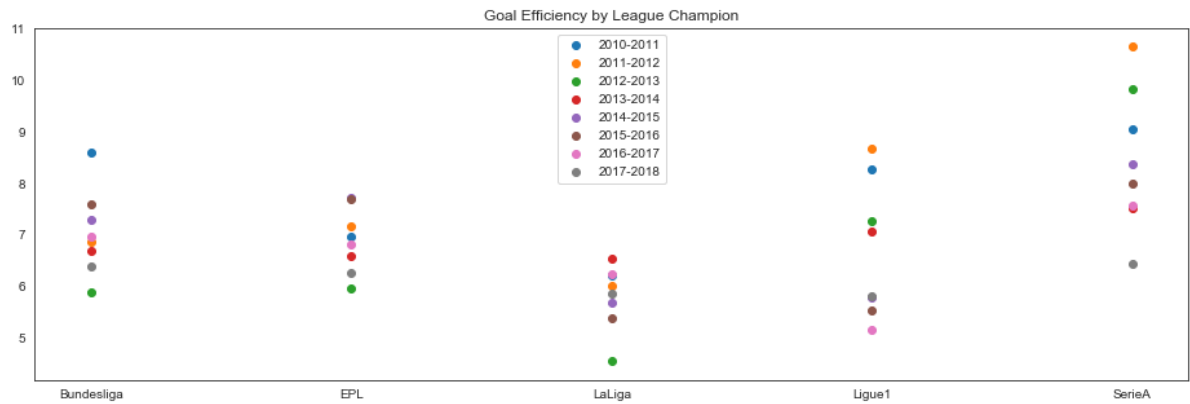
The trend is similar to the trend of goal except Serie A. The Serie A champion tend to attempt more. Maybe we should consider the goal per shot as a key feature to compare with leagues.

```
Out[204]: League
Bundesliga    657.500
EPL           590.500
LaLiga       602.250
Ligue1       558.750
SerieA       618.875
dtype: float64
```

```
Out[205]: League
Bundesliga    400.716049
EPL           448.405556
LaLiga       433.266667
Ligue1       418.961111
SerieA       457.516667
dtype: float64
```

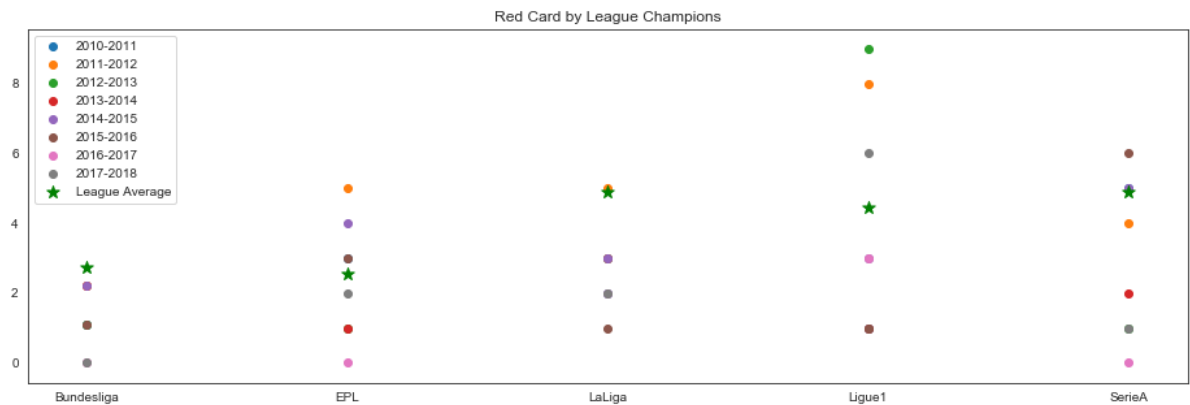
Similar to the number of goal analysis, the total number of shots for champion's team is much higher than the league average.

<Figure size 432x288 with 0 Axes>



As we expected, Serie A has the lowest Goal coefficient

<Figure size 432x288 with 0 Axes>



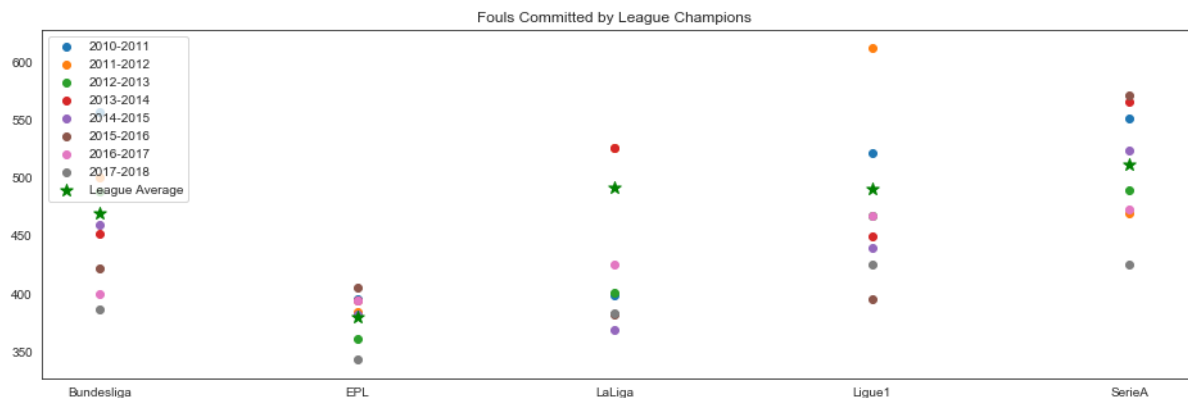
```
Out[209]: League
Bundesliga    1.250
EPL           2.375
LaLiga        2.625
Ligue1        4.000
SerieA        3.000
dtype: float64
```

```
Out[210]: League
Bundesliga    2.734568
EPL           2.550000
LaLiga        4.911111
Ligue1        4.430556
SerieA        4.884259
dtype: float64
```

The red card received by champion team is also lower than league average.

The number of red card by champions are low except Ligue 1.

<Figure size 432x288 with 0 Axes>



```
mean = Champs.groupby('League').mean() Allmean = ww.groupby('League').mean()
mean['HomeTeam_FoulsCommitted']+mean['AwayTeam_FoulsCommitted']
```

```
Out[212]: League
Bundesliga    469.067901
EPL           379.655556
LaLiga        491.800000
Ligue1        490.271528
SerieA        511.836111
dtype: float64
```

Except Serie A, all other 4 league's champion's fouls committed is under average number of fouls committed by all teams.

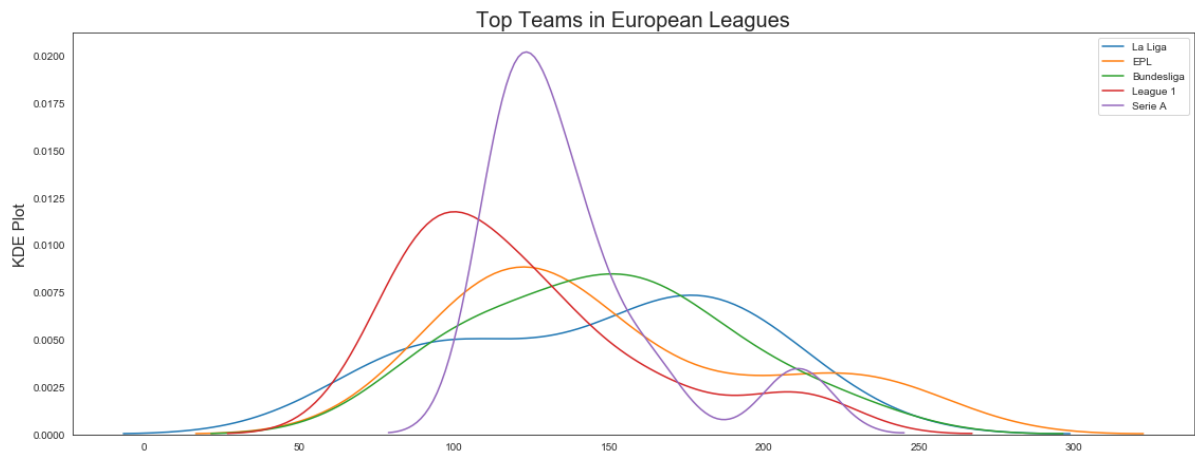
Competitiveness of Leagues

Which European league is the most competitive ? Lets try to find an answer to this question

The gap between the top team and 20th team in a league may be too much (and also might be influenced by outliers). So we'll limit our analysis to top 10 teams of every league

```
C:\Users\ksaha\AppData\Local\Continuum\anaconda3\lib\site-packages\scipy\stats\stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.
    return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```

<Figure size 432x288 with 0 Axes>



Leagues that are normally distributed are less competitive because the best and worst teams are equidistant from the mean

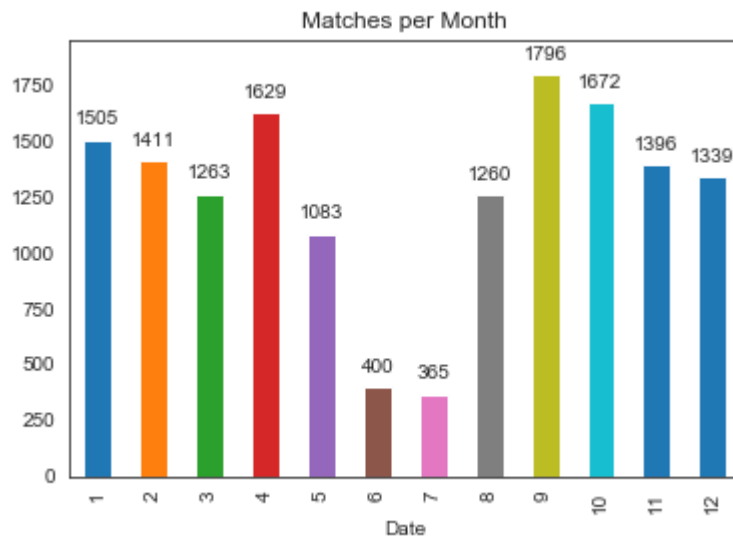
Left skewed distribution means only a few handful teams are close to the top whereas right skewed distribution means there's lot of competition at the top

From the plot it seems that La Liga and Bundesliga are the most competitive

Time Series Analysis

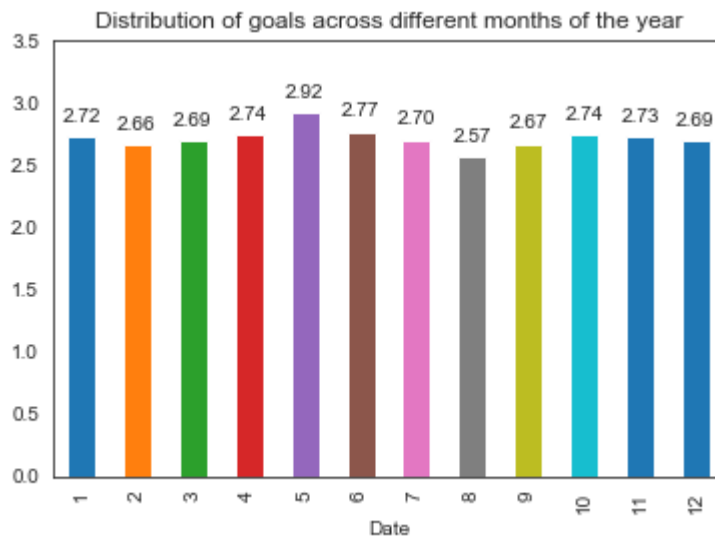
Lets convert the data into a time series first

During which months of the year are most games played ?



Very few matches are played during the months of June and July. Most of the matches are played during Fall (September-October)

Is there a seasonal trend in goal scoring rate ? Does goal scoring rate dip during winter months ?



Clearly there is no significant dip in goal scoring rate during Winter months. Most goals are scored in May/June, towards the end of the year

Predictive Analysis

Statistical studies on goal scoring pattern in football matches often suggest that goals in a football match follow Poisson distribution pattern. We'll continue with that assumption and use the the mean expected number of goals for a match to form a probability distribution and express the number of goals as a function of average rate of scoring goals

We'll use 2010-2011 to 2017-2018 data to train our model and 2018-2019 data to test our model

We'll create two distributions - one for home and other for away

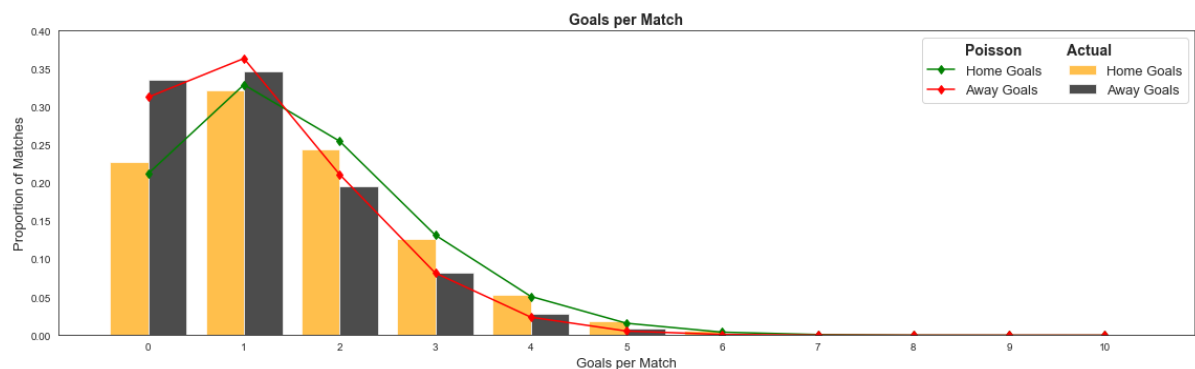
First lets look at the data distribution of home and away goals

```
Full Time Home-Team Goals (Unique Values) : : [ 3  1  0  6  2  4  5  7  8  9
10]
Full Time Away-Team Goals (Unique Values) : : [0  2  4  1  3  6  5  7  8  9]
```

```
C:\Users\ksaha\AppData\Local\Continuum\anaconda3\lib\site-packages\matplotlib
\axes\_axes.py:6571: UserWarning: The 'normed' kwarg is deprecated, and has b
een replaced by the 'density' kwarg.
```

```
warnings.warn("The 'normed' kwarg is deprecated, and has been ")
```

```
<Figure size 432x288 with 0 Axes>
```



Now lets build a regression model based on Poisson distribution.

Our model formula will be :

Home/Away Goals = Opponent Team + Away/Home Goals

Since teams across leagues dont play each other (We dont have Champions league data), so we will build a separate model per league

Out[221]:

Generalized Linear Model Regression Results

| | | | |
|------------------------|------------------|--------------------------|-----------|
| Dep. Variable: | goals | No. Observations: | 6080 |
| Model: | GLM | Df Residuals: | 6010 |
| Model Family: | Poisson | Df Model: | 69 |
| Link Function: | log | Scale: | 1.0000 |
| Method: | IRLS | Log-Likelihood: | -8779.0 |
| Date: | Fri, 07 Dec 2018 | Deviance: | 6823.0 |
| Time: | 15:23:42 | Pearson chi2: | 5.93e+03 |
| No. Iterations: | 5 | Covariance Type: | nonrobust |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|------------------------|---------|---------|--------|-------|--------|--------|
| Intercept | 0.2893 | 0.071 | 4.090 | 0.000 | 0.151 | 0.428 |
| team[T.Aston Villa] | -0.6235 | 0.078 | -7.962 | 0.000 | -0.777 | -0.470 |
| team[T.Birmingham] | -0.7025 | 0.170 | -4.129 | 0.000 | -1.036 | -0.369 |
| team[T.Blackburn] | -0.4477 | 0.112 | -4.005 | 0.000 | -0.667 | -0.229 |
| team[T.Blackpool] | -0.2857 | 0.142 | -2.016 | 0.044 | -0.564 | -0.008 |
| team[T.Bolton] | -0.4078 | 0.110 | -3.714 | 0.000 | -0.623 | -0.193 |
| team[T.Bournemouth] | -0.3393 | 0.093 | -3.637 | 0.000 | -0.522 | -0.156 |
| team[T.Brighton] | -0.6946 | 0.177 | -3.927 | 0.000 | -1.041 | -0.348 |
| team[T.Burnley] | -0.6942 | 0.107 | -6.471 | 0.000 | -0.904 | -0.484 |
| team[T.Cardiff] | -0.7750 | 0.182 | -4.263 | 0.000 | -1.131 | -0.419 |
| team[T.Chelsea] | -0.0294 | 0.060 | -0.494 | 0.621 | -0.146 | 0.087 |
| team[T.Crystal Palace] | -0.4800 | 0.080 | -5.975 | 0.000 | -0.637 | -0.323 |
| team[T.Everton] | -0.2836 | 0.064 | -4.439 | 0.000 | -0.409 | -0.158 |
| team[T.Fulham] | -0.4400 | 0.084 | -5.209 | 0.000 | -0.606 | -0.274 |
| team[T.Huddersfield] | -0.8850 | 0.194 | -4.564 | 0.000 | -1.265 | -0.505 |
| team[T.Hull] | -0.6475 | 0.105 | -6.162 | 0.000 | -0.853 | -0.442 |
| team[T.Leicester] | -0.2316 | 0.080 | -2.900 | 0.004 | -0.388 | -0.075 |
| team[T.Liverpool] | -0.0297 | 0.060 | -0.498 | 0.618 | -0.147 | 0.087 |
| team[T.Man City] | 0.1357 | 0.057 | 2.374 | 0.018 | 0.024 | 0.248 |
| team[T.Man United] | -0.0477 | 0.060 | -0.798 | 0.425 | -0.165 | 0.069 |
| team[T.Middlesbrough] | -0.9278 | 0.197 | -4.706 | 0.000 | -1.314 | -0.541 |
| team[T.Newcastle] | -0.4297 | 0.070 | -6.167 | 0.000 | -0.566 | -0.293 |
| team[T.Norwich] | -0.5749 | 0.090 | -6.420 | 0.000 | -0.750 | -0.399 |
| team[T.QPR] | -0.6141 | 0.102 | -6.001 | 0.000 | -0.815 | -0.414 |
| team[T.Reading] | -0.4968 | 0.158 | -3.138 | 0.002 | -0.807 | -0.186 |
| team[T.Southampton] | -0.3578 | 0.072 | -4.975 | 0.000 | -0.499 | -0.217 |

| | | | | | | |
|----------------------------|---------|-------|--------|-------|--------|--------|
| team[T.Stoke] | -0.5547 | 0.069 | -7.986 | 0.000 | -0.691 | -0.419 |
| team[T.Sunderland] | -0.5745 | 0.073 | -7.869 | 0.000 | -0.718 | -0.431 |
| team[T.Swansea] | -0.4740 | 0.071 | -6.684 | 0.000 | -0.613 | -0.335 |
| team[T.Tottenham] | -0.0805 | 0.060 | -1.333 | 0.182 | -0.199 | 0.038 |
| team[T.Watford] | -0.5008 | 0.099 | -5.041 | 0.000 | -0.696 | -0.306 |
| team[T.West Brom] | -0.5009 | 0.068 | -7.327 | 0.000 | -0.635 | -0.367 |
| team[T.West Ham] | -0.3907 | 0.069 | -5.656 | 0.000 | -0.526 | -0.255 |
| team[T.Wigan] | -0.5283 | 0.098 | -5.402 | 0.000 | -0.720 | -0.337 |
| team[T.Wolves] | -0.5315 | 0.116 | -4.577 | 0.000 | -0.759 | -0.304 |
| opponent[T.Aston Villa] | 0.3613 | 0.075 | 4.806 | 0.000 | 0.214 | 0.509 |
| opponent[T.Birmingham] | 0.2793 | 0.143 | 1.958 | 0.050 | -0.000 | 0.559 |
| opponent[T.Blackburn] | 0.4629 | 0.102 | 4.550 | 0.000 | 0.263 | 0.662 |
| opponent[T.Blackpool] | 0.5938 | 0.126 | 4.708 | 0.000 | 0.347 | 0.841 |
| opponent[T.Bolton] | 0.4351 | 0.103 | 4.232 | 0.000 | 0.234 | 0.637 |
| opponent[T.Bournemouth] | 0.4052 | 0.090 | 4.491 | 0.000 | 0.228 | 0.582 |
| opponent[T.Brighton] | 0.2087 | 0.147 | 1.421 | 0.155 | -0.079 | 0.497 |
| opponent[T.Burnley] | 0.1150 | 0.099 | 1.161 | 0.246 | -0.079 | 0.309 |
| opponent[T.Cardiff] | 0.5376 | 0.129 | 4.182 | 0.000 | 0.286 | 0.790 |
| opponent[T.Chelsea] | -0.1151 | 0.079 | -1.449 | 0.147 | -0.271 | 0.041 |
| opponent[T.Crystal Palace] | 0.2116 | 0.082 | 2.579 | 0.010 | 0.051 | 0.372 |
| opponent[T.Everton] | 0.0782 | 0.075 | 1.037 | 0.300 | -0.070 | 0.226 |
| opponent[T.Fulham] | 0.3293 | 0.085 | 3.883 | 0.000 | 0.163 | 0.495 |
| opponent[T.Huddersfield] | 0.2741 | 0.142 | 1.924 | 0.054 | -0.005 | 0.553 |
| opponent[T.Hull] | 0.3467 | 0.092 | 3.776 | 0.000 | 0.167 | 0.527 |
| opponent[T.Leicester] | 0.2185 | 0.088 | 2.493 | 0.013 | 0.047 | 0.390 |
| opponent[T.Liverpool] | 0.0499 | 0.076 | 0.655 | 0.513 | -0.099 | 0.199 |
| opponent[T.Man City] | -0.1817 | 0.081 | -2.240 | 0.025 | -0.341 | -0.023 |
| opponent[T.Man United] | -0.1710 | 0.081 | -2.122 | 0.034 | -0.329 | -0.013 |
| opponent[T.Middlesbrough] | 0.1884 | 0.148 | 1.273 | 0.203 | -0.102 | 0.478 |
| opponent[T.Newcastle] | 0.3040 | 0.074 | 4.128 | 0.000 | 0.160 | 0.448 |
| opponent[T.Norwich] | 0.3760 | 0.083 | 4.513 | 0.000 | 0.213 | 0.539 |
| opponent[T.QPR] | 0.4257 | 0.090 | 4.754 | 0.000 | 0.250 | 0.601 |
| opponent[T.Reading] | 0.5248 | 0.129 | 4.059 | 0.000 | 0.271 | 0.778 |
| opponent[T.Southampton] | 0.0938 | 0.081 | 1.163 | 0.245 | -0.064 | 0.252 |
| opponent[T.Stoke] | 0.1942 | 0.073 | 2.654 | 0.008 | 0.051 | 0.338 |
| opponent[T.Sunderland] | 0.2739 | 0.074 | 3.699 | 0.000 | 0.129 | 0.419 |
| opponent[T.Swansea] | 0.2348 | 0.075 | 3.139 | 0.002 | 0.088 | 0.381 |

| | | | | | | |
|------------------------------|---------|-------|--------|-------|--------|-------|
| opponent[T.Tottenham] | -0.0146 | 0.077 | -0.189 | 0.850 | -0.166 | 0.137 |
| opponent[T.Watford] | 0.3286 | 0.092 | 3.563 | 0.000 | 0.148 | 0.509 |
| opponent[T.West Brom] | 0.2495 | 0.072 | 3.450 | 0.001 | 0.108 | 0.391 |
| opponent[T.West Ham] | 0.2900 | 0.074 | 3.925 | 0.000 | 0.145 | 0.435 |
| opponent[T.Wigan] | 0.4122 | 0.090 | 4.573 | 0.000 | 0.236 | 0.589 |
| opponent[T.Wolves] | 0.5365 | 0.099 | 5.416 | 0.000 | 0.342 | 0.731 |
| home | 0.2724 | 0.022 | 12.312 | 0.000 | 0.229 | 0.316 |

Out[222]:

Generalized Linear Model Regression Results

| | | | |
|------------------------|------------------|--------------------------|-----------|
| Dep. Variable: | goals | No. Observations: | 6080 |
| Model: | GLM | Df Residuals: | 6016 |
| Model Family: | Poisson | Df Model: | 63 |
| Link Function: | log | Scale: | 1.0000 |
| Method: | IRLS | Log-Likelihood: | -8695.6 |
| Date: | Fri, 07 Dec 2018 | Deviance: | 6710.9 |
| Time: | 15:23:42 | Pearson chi2: | 5.81e+03 |
| No. Iterations: | 5 | Covariance Type: | nonrobust |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|----------------------------|---------|---------|--------|-------|--------|--------|
| Intercept | -0.2524 | 0.153 | -1.647 | 0.100 | -0.553 | 0.048 |
| team[T.Almeria] | -0.0743 | 0.146 | -0.510 | 0.610 | -0.360 | 0.211 |
| team[T.Ath Bilbao] | 0.2167 | 0.122 | 1.779 | 0.075 | -0.022 | 0.456 |
| team[T.Ath Madrid] | 0.4233 | 0.120 | 3.533 | 0.000 | 0.188 | 0.658 |
| team[T.Barcelona] | 0.9370 | 0.117 | 8.041 | 0.000 | 0.709 | 1.165 |
| team[T.Betis] | 0.1141 | 0.127 | 0.901 | 0.368 | -0.134 | 0.362 |
| team[T.Celta] | 0.1826 | 0.126 | 1.453 | 0.146 | -0.064 | 0.429 |
| team[T.Cordoba] | -0.6250 | 0.241 | -2.596 | 0.009 | -1.097 | -0.153 |
| team[T.Eibar] | 0.1128 | 0.134 | 0.844 | 0.398 | -0.149 | 0.375 |
| team[T.Elche] | -0.2459 | 0.167 | -1.473 | 0.141 | -0.573 | 0.081 |
| team[T.Espanol] | 0.0527 | 0.124 | 0.427 | 0.670 | -0.190 | 0.295 |
| team[T.Getafe] | -0.0362 | 0.127 | -0.286 | 0.775 | -0.284 | 0.212 |
| team[T.Girona] | 0.2321 | 0.180 | 1.290 | 0.197 | -0.121 | 0.585 |
| team[T.Granada] | -0.1658 | 0.131 | -1.264 | 0.206 | -0.423 | 0.091 |
| team[T.Hercules] | -0.1318 | 0.201 | -0.657 | 0.511 | -0.525 | 0.262 |
| team[T.La Coruna] | -0.0244 | 0.129 | -0.190 | 0.850 | -0.277 | 0.228 |
| team[T.Las Palmas] | 0.0137 | 0.143 | 0.095 | 0.924 | -0.268 | 0.295 |
| team[T.Leganés] | -0.1397 | 0.163 | -0.855 | 0.392 | -0.460 | 0.180 |
| team[T.Levante] | -0.0133 | 0.126 | -0.106 | 0.916 | -0.261 | 0.234 |
| team[T.Malaga] | 0.0652 | 0.123 | 0.528 | 0.598 | -0.177 | 0.307 |
| team[T.Mallorca] | 0.0124 | 0.143 | 0.086 | 0.931 | -0.268 | 0.292 |
| team[T.Osasuna] | -0.0568 | 0.133 | -0.428 | 0.668 | -0.317 | 0.203 |
| team[T.Real Madrid] | 0.9417 | 0.117 | 8.080 | 0.000 | 0.713 | 1.170 |
| team[T.Santander] | -0.1777 | 0.164 | -1.081 | 0.280 | -0.500 | 0.144 |
| team[T.Sevilla] | 0.3670 | 0.120 | 3.047 | 0.002 | 0.131 | 0.603 |
| team[T.Sociedad] | 0.2905 | 0.121 | 2.397 | 0.017 | 0.053 | 0.528 |

| | | | | | | |
|-------------------------|---------|-------|--------|-------|--------|--------|
| team[T.Sp Gijon] | -0.0292 | 0.137 | -0.214 | 0.831 | -0.297 | 0.239 |
| team[T.Valencia] | 0.3643 | 0.120 | 3.025 | 0.002 | 0.128 | 0.600 |
| team[T.Valladolid] | 0.0434 | 0.155 | 0.280 | 0.779 | -0.260 | 0.347 |
| team[T.Vallecano] | 0.1883 | 0.128 | 1.467 | 0.142 | -0.063 | 0.440 |
| team[T.Villarreal] | 0.2064 | 0.123 | 1.674 | 0.094 | -0.035 | 0.448 |
| team[T.Zaragoza] | -0.0961 | 0.146 | -0.658 | 0.511 | -0.382 | 0.190 |
| opponent[T.Almeria] | 0.4077 | 0.125 | 3.252 | 0.001 | 0.162 | 0.653 |
| opponent[T.Ath Bilbao] | 0.0676 | 0.116 | 0.585 | 0.559 | -0.159 | 0.294 |
| opponent[T.Ath Madrid] | -0.3547 | 0.122 | -2.918 | 0.004 | -0.593 | -0.116 |
| opponent[T.Barcelona] | -0.3638 | 0.122 | -2.971 | 0.003 | -0.604 | -0.124 |
| opponent[T.Betis] | 0.2854 | 0.116 | 2.454 | 0.014 | 0.057 | 0.513 |
| opponent[T.Celta] | 0.2081 | 0.117 | 1.774 | 0.076 | -0.022 | 0.438 |
| opponent[T.Cordoba] | 0.3913 | 0.160 | 2.447 | 0.014 | 0.078 | 0.705 |
| opponent[T.Eibar] | 0.1653 | 0.124 | 1.332 | 0.183 | -0.078 | 0.408 |
| opponent[T.Elche] | 0.1974 | 0.141 | 1.404 | 0.160 | -0.078 | 0.473 |
| opponent[T.Espanol] | 0.1624 | 0.115 | 1.418 | 0.156 | -0.062 | 0.387 |
| opponent[T.Getafe] | 0.1826 | 0.116 | 1.577 | 0.115 | -0.044 | 0.409 |
| opponent[T.Girona] | 0.2396 | 0.167 | 1.438 | 0.150 | -0.087 | 0.566 |
| opponent[T.Granada] | 0.3182 | 0.116 | 2.746 | 0.006 | 0.091 | 0.545 |
| opponent[T.Hercules] | 0.2860 | 0.166 | 1.723 | 0.085 | -0.039 | 0.611 |
| opponent[T.La Coruna] | 0.3070 | 0.116 | 2.646 | 0.008 | 0.080 | 0.534 |
| opponent[T.Las Palmas] | 0.3642 | 0.126 | 2.901 | 0.004 | 0.118 | 0.610 |
| opponent[T.Leganés] | 0.1256 | 0.142 | 0.883 | 0.377 | -0.153 | 0.404 |
| opponent[T.Levante] | 0.2116 | 0.115 | 1.833 | 0.067 | -0.015 | 0.438 |
| opponent[T.Málaga] | 0.1275 | 0.115 | 1.110 | 0.267 | -0.098 | 0.353 |
| opponent[T.Mallorca] | 0.2443 | 0.129 | 1.896 | 0.058 | -0.008 | 0.497 |
| opponent[T.Osasuna] | 0.3126 | 0.118 | 2.641 | 0.008 | 0.081 | 0.545 |
| opponent[T.Real Madrid] | -0.1293 | 0.119 | -1.088 | 0.277 | -0.362 | 0.104 |
| opponent[T.Santander] | 0.2678 | 0.139 | 1.929 | 0.054 | -0.004 | 0.540 |
| opponent[T.Sevilla] | 0.1431 | 0.115 | 1.245 | 0.213 | -0.082 | 0.368 |
| opponent[T.Sociedad] | 0.1787 | 0.115 | 1.561 | 0.119 | -0.046 | 0.403 |
| opponent[T.Sp Gijon] | 0.2893 | 0.122 | 2.371 | 0.018 | 0.050 | 0.528 |
| opponent[T.Valencia] | 0.0472 | 0.116 | 0.407 | 0.684 | -0.180 | 0.274 |
| opponent[T.Valladolid] | 0.2514 | 0.139 | 1.809 | 0.071 | -0.021 | 0.524 |
| opponent[T.Vallecano] | 0.4594 | 0.117 | 3.940 | 0.000 | 0.231 | 0.688 |
| opponent[T.Villarreal] | -0.0725 | 0.119 | -0.609 | 0.543 | -0.306 | 0.161 |
| opponent[T.Zaragoza] | 0.2514 | 0.129 | 1.955 | 0.051 | -0.001 | 0.503 |

| | | | | | | |
|-------------|--------|-------|--------|-------|-------|-------|
| home | 0.3498 | 0.022 | 15.807 | 0.000 | 0.306 | 0.393 |
|-------------|--------|-------|--------|-------|-------|-------|

Out[223]:

Generalized Linear Model Regression Results

| | | | |
|------------------------|------------------|--------------------------|-----------|
| Dep. Variable: | goals | No. Observations: | 6080 |
| Model: | GLM | Df Residuals: | 6012 |
| Model Family: | Poisson | Df Model: | 67 |
| Link Function: | log | Scale: | 1.0000 |
| Method: | IRLS | Log-Likelihood: | -8656.9 |
| Date: | Fri, 07 Dec 2018 | Deviance: | 6684.9 |
| Time: | 15:23:42 | Pearson chi2: | 5.81e+03 |
| No. Iterations: | 5 | Covariance Type: | nonrobust |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|---------------------------|---------|---------|--------|-------|--------|--------|
| Intercept | -0.0226 | 0.080 | -0.282 | 0.778 | -0.179 | 0.134 |
| team[T.Bari] | -0.5114 | 0.201 | -2.548 | 0.011 | -0.905 | -0.118 |
| team[T.Benevento] | -0.3105 | 0.183 | -1.696 | 0.090 | -0.669 | 0.048 |
| team[T.Bologna] | -0.1953 | 0.083 | -2.345 | 0.019 | -0.358 | -0.032 |
| team[T.Brescia] | -0.2847 | 0.181 | -1.575 | 0.115 | -0.639 | 0.070 |
| team[T.Cagliari] | -0.0748 | 0.081 | -0.925 | 0.355 | -0.233 | 0.084 |
| team[T.Carpi] | -0.2091 | 0.174 | -1.202 | 0.229 | -0.550 | 0.132 |
| team[T.Catania] | -0.0559 | 0.095 | -0.589 | 0.556 | -0.242 | 0.130 |
| team[T.Cesena] | -0.3172 | 0.116 | -2.743 | 0.006 | -0.544 | -0.091 |
| team[T.Chievo] | -0.2186 | 0.081 | -2.705 | 0.007 | -0.377 | -0.060 |
| team[T.Crotone] | -0.2121 | 0.129 | -1.642 | 0.101 | -0.465 | 0.041 |
| team[T.Empoli] | -0.1741 | 0.109 | -1.600 | 0.110 | -0.387 | 0.039 |
| team[T.Fiorentina] | 0.2283 | 0.073 | 3.136 | 0.002 | 0.086 | 0.371 |
| team[T.Frosinone] | -0.2451 | 0.178 | -1.375 | 0.169 | -0.595 | 0.104 |
| team[T.Genoa] | -0.0352 | 0.077 | -0.455 | 0.649 | -0.187 | 0.116 |
| team[T.Inter] | 0.2905 | 0.072 | 4.041 | 0.000 | 0.150 | 0.431 |
| team[T.Juventus] | 0.4483 | 0.070 | 6.448 | 0.000 | 0.312 | 0.585 |
| team[T.Lazio] | 0.3145 | 0.072 | 4.395 | 0.000 | 0.174 | 0.455 |
| team[T.Lecce] | -0.0401 | 0.122 | -0.329 | 0.742 | -0.279 | 0.199 |
| team[T.Livorno] | -0.1300 | 0.170 | -0.766 | 0.444 | -0.463 | 0.203 |
| team[T.Milan] | 0.2655 | 0.072 | 3.679 | 0.000 | 0.124 | 0.407 |
| team[T.Napoli] | 0.4791 | 0.069 | 6.910 | 0.000 | 0.343 | 0.615 |
| team[T.Novara] | -0.2427 | 0.178 | -1.361 | 0.173 | -0.592 | 0.107 |
| team[T.Palermo] | -0.0063 | 0.083 | -0.076 | 0.939 | -0.169 | 0.156 |
| team[T.Parma] | 0.0122 | 0.087 | 0.141 | 0.888 | -0.158 | 0.182 |
| team[T.Pescara] | -0.3227 | 0.137 | -2.355 | 0.019 | -0.591 | -0.054 |

| | | | | | | |
|------------------------|---------|-------|--------|-------|--------|--------|
| team[T.Roma] | 0.3999 | 0.070 | 5.686 | 0.000 | 0.262 | 0.538 |
| team[T.Sampdoria] | 0.0193 | 0.079 | 0.245 | 0.806 | -0.135 | 0.174 |
| team[T.Sassuolo] | 0.0014 | 0.087 | 0.017 | 0.987 | -0.169 | 0.171 |
| team[T.Siena] | -0.1118 | 0.125 | -0.897 | 0.370 | -0.356 | 0.132 |
| team[T.Spal] | -0.1687 | 0.170 | -0.993 | 0.321 | -0.502 | 0.164 |
| team[T.Torino] | 0.1826 | 0.079 | 2.324 | 0.020 | 0.029 | 0.337 |
| team[T.Udinese] | 0.0797 | 0.075 | 1.060 | 0.289 | -0.068 | 0.227 |
| team[T.Verona] | -0.0290 | 0.094 | -0.308 | 0.758 | -0.214 | 0.155 |
| opponent[T.Bari] | 0.1740 | 0.145 | 1.202 | 0.229 | -0.110 | 0.458 |
| opponent[T.Benevento] | 0.5590 | 0.122 | 4.569 | 0.000 | 0.319 | 0.799 |
| opponent[T.Bologna] | 0.0729 | 0.076 | 0.958 | 0.338 | -0.076 | 0.222 |
| opponent[T.Brescia] | 0.1072 | 0.149 | 0.718 | 0.473 | -0.186 | 0.400 |
| opponent[T.Cagliari] | 0.2062 | 0.074 | 2.794 | 0.005 | 0.062 | 0.351 |
| opponent[T.Carpi] | 0.1696 | 0.144 | 1.182 | 0.237 | -0.112 | 0.451 |
| opponent[T.Catania] | 0.1291 | 0.087 | 1.476 | 0.140 | -0.042 | 0.301 |
| opponent[T.Cesena] | 0.2461 | 0.092 | 2.670 | 0.008 | 0.065 | 0.427 |
| opponent[T.Chievo] | 0.0341 | 0.074 | 0.459 | 0.646 | -0.112 | 0.180 |
| opponent[T.Crotone] | 0.2587 | 0.105 | 2.456 | 0.014 | 0.052 | 0.465 |
| opponent[T.Empoli] | 0.1146 | 0.096 | 1.195 | 0.232 | -0.073 | 0.302 |
| opponent[T.Fiorentina] | -0.0258 | 0.076 | -0.340 | 0.734 | -0.174 | 0.123 |
| opponent[T.Frosinone] | 0.4560 | 0.127 | 3.580 | 0.000 | 0.206 | 0.706 |
| opponent[T.Genoa] | 0.0980 | 0.073 | 1.336 | 0.182 | -0.046 | 0.242 |
| opponent[T.Inter] | -0.0440 | 0.076 | -0.577 | 0.564 | -0.193 | 0.105 |
| opponent[T.Juventus] | -0.5711 | 0.088 | -6.469 | 0.000 | -0.744 | -0.398 |
| opponent[T.Lazio] | -0.0041 | 0.075 | -0.054 | 0.957 | -0.152 | 0.144 |
| opponent[T.Lecce] | 0.2695 | 0.106 | 2.540 | 0.011 | 0.062 | 0.477 |
| opponent[T.Livorno] | 0.4579 | 0.127 | 3.618 | 0.000 | 0.210 | 0.706 |
| opponent[T.Milan] | -0.1423 | 0.078 | -1.824 | 0.068 | -0.295 | 0.011 |
| opponent[T.Napoli] | -0.1617 | 0.079 | -2.054 | 0.040 | -0.316 | -0.007 |
| opponent[T.Novara] | 0.3172 | 0.136 | 2.335 | 0.020 | 0.051 | 0.584 |
| opponent[T.Palermo] | 0.2793 | 0.075 | 3.709 | 0.000 | 0.132 | 0.427 |
| opponent[T.Parma] | 0.1173 | 0.082 | 1.426 | 0.154 | -0.044 | 0.278 |
| opponent[T.Pescara] | 0.5355 | 0.095 | 5.620 | 0.000 | 0.349 | 0.722 |
| opponent[T.Roma] | -0.1332 | 0.078 | -1.708 | 0.088 | -0.286 | 0.020 |
| opponent[T.Sampdoria] | 0.1320 | 0.075 | 1.758 | 0.079 | -0.015 | 0.279 |
| opponent[T.Sassuolo] | 0.1959 | 0.080 | 2.439 | 0.015 | 0.038 | 0.353 |
| opponent[T.Siena] | 0.0688 | 0.113 | 0.607 | 0.544 | -0.153 | 0.291 |

| | | | | | | |
|----------------------------|--------|-------|--------|-------|--------|-------|
| opponent[T.Spal] | 0.2110 | 0.141 | 1.491 | 0.136 | -0.066 | 0.488 |
| opponent[T.Torino] | 0.1020 | 0.079 | 1.297 | 0.195 | -0.052 | 0.256 |
| opponent[T.Udinese] | 0.0916 | 0.074 | 1.245 | 0.213 | -0.053 | 0.236 |
| opponent[T.Verona] | 0.3553 | 0.082 | 4.353 | 0.000 | 0.195 | 0.515 |
| home | 0.2600 | 0.022 | 11.606 | 0.000 | 0.216 | 0.304 |

Out[224]:

Generalized Linear Model Regression Results

| | | | |
|------------------------|------------------|--------------------------|-----------|
| Dep. Variable: | goals | No. Observations: | 4896 |
| Model: | GLM | Df Residuals: | 4842 |
| Model Family: | Poisson | Df Model: | 53 |
| Link Function: | log | Scale: | 1.0000 |
| Method: | IRLS | Log-Likelihood: | -7244.0 |
| Date: | Fri, 07 Dec 2018 | Deviance: | 5515.4 |
| Time: | 15:23:43 | Pearson chi2: | 4.78e+03 |
| No. Iterations: | 5 | Covariance Type: | nonrobust |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|----------------------------|---------|---------|--------|-------|--------|--------|
| Intercept | -0.0023 | 0.083 | -0.027 | 0.978 | -0.165 | 0.160 |
| team[T.Bayern Munich] | 0.7428 | 0.071 | 10.457 | 0.000 | 0.604 | 0.882 |
| team[T.Braunschweig] | -0.3079 | 0.195 | -1.577 | 0.115 | -0.691 | 0.075 |
| team[T.Darmstadt] | -0.1688 | 0.137 | -1.231 | 0.218 | -0.437 | 0.100 |
| team[T.Dortmund] | 0.5704 | 0.073 | 7.802 | 0.000 | 0.427 | 0.714 |
| team[T.Ein Frankfurt] | 0.0413 | 0.084 | 0.493 | 0.622 | -0.123 | 0.206 |
| team[T.FC Koln] | 0.0267 | 0.088 | 0.304 | 0.761 | -0.145 | 0.199 |
| team[T.Fortuna Dusseldorf] | -0.0237 | 0.171 | -0.138 | 0.890 | -0.359 | 0.312 |
| team[T.Freiburg] | 0.0202 | 0.084 | 0.239 | 0.811 | -0.145 | 0.186 |
| team[T.Greuther Furth] | -0.4266 | 0.205 | -2.078 | 0.038 | -0.829 | -0.024 |
| team[T.Hamburg] | -0.0519 | 0.083 | -0.624 | 0.533 | -0.215 | 0.111 |
| team[T.Hannover] | 0.1108 | 0.083 | 1.341 | 0.180 | -0.051 | 0.273 |
| team[T.Hertha] | 0.0156 | 0.088 | 0.178 | 0.859 | -0.157 | 0.188 |
| team[T.Hoffenheim] | 0.2867 | 0.077 | 3.711 | 0.000 | 0.135 | 0.438 |
| team[T.Ingolstadt] | -0.1344 | 0.135 | -0.998 | 0.318 | -0.398 | 0.129 |
| team[T.Kaiserslautern] | -0.1091 | 0.132 | -0.824 | 0.410 | -0.369 | 0.150 |
| team[T.Leverkusen] | 0.3797 | 0.076 | 5.017 | 0.000 | 0.231 | 0.528 |
| team[T.M'gladbach] | 0.2521 | 0.078 | 3.248 | 0.001 | 0.100 | 0.404 |
| team[T.Mainz] | 0.1348 | 0.080 | 1.693 | 0.090 | -0.021 | 0.291 |
| team[T.Nurnberg] | 0.0053 | 0.099 | 0.053 | 0.957 | -0.189 | 0.200 |
| team[T.Paderborn] | -0.2436 | 0.189 | -1.286 | 0.198 | -0.615 | 0.128 |
| team[T.RB Leipzig] | 0.4444 | 0.108 | 4.100 | 0.000 | 0.232 | 0.657 |
| team[T.Schalke 04] | 0.2768 | 0.077 | 3.586 | 0.000 | 0.126 | 0.428 |
| team[T.St Pauli] | -0.1310 | 0.179 | -0.730 | 0.466 | -0.483 | 0.221 |
| team[T.Stuttgart] | 0.1931 | 0.081 | 2.382 | 0.017 | 0.034 | 0.352 |
| team[T.Werder Bremen] | 0.2035 | 0.079 | 2.585 | 0.010 | 0.049 | 0.358 |

| | | | | | | |
|--------------------------------|---------|-------|--------|-------|-----------|--------|
| team[T.Wolfsburg] | 0.1979 | 0.079 | 2.518 | 0.012 | 0.044 | 0.352 |
| opponent[T.Bayern Munich] | -0.6688 | 0.091 | -7.344 | 0.000 | -0.847 | -0.490 |
| opponent[T.Braunschweig] | 0.2096 | 0.140 | 1.495 | 0.135 | -0.065 | 0.484 |
| opponent[T.Darmstadt] | 0.1757 | 0.108 | 1.630 | 0.103 | -0.035 | 0.387 |
| opponent[T.Dortmund] | -0.2519 | 0.080 | -3.144 | 0.002 | -0.409 | -0.095 |
| opponent[T.Ein Frankfurt] | 0.0459 | 0.076 | 0.604 | 0.546 | -0.103 | 0.195 |
| opponent[T.FC Koln] | 0.1289 | 0.077 | 1.666 | 0.096 | -0.023 | 0.281 |
| opponent[T.Fortuna Dusseldorf] | 0.1755 | 0.143 | 1.224 | 0.221 | -0.105 | 0.457 |
| opponent[T.Freiburg] | 0.1015 | 0.075 | 1.352 | 0.176 | -0.046 | 0.249 |
| opponent[T.Greuther Furth] | 0.2114 | 0.140 | 1.507 | 0.132 | -0.064 | 0.486 |
| opponent[T.Hamburg] | 0.1413 | 0.072 | 1.959 | 0.050 | -3.38e-05 | 0.283 |
| opponent[T.Hannover] | 0.1289 | 0.075 | 1.725 | 0.084 | -0.018 | 0.275 |
| opponent[T.Hertha] | 0.0271 | 0.079 | 0.341 | 0.733 | -0.129 | 0.183 |
| opponent[T.Hoffenheim] | 0.1161 | 0.073 | 1.595 | 0.111 | -0.027 | 0.259 |
| opponent[T.Ingolstadt] | 0.0186 | 0.114 | 0.162 | 0.871 | -0.206 | 0.243 |
| opponent[T.Kaiserslautern] | 0.0790 | 0.112 | 0.707 | 0.480 | -0.140 | 0.298 |
| opponent[T.Leverkusen] | -0.0961 | 0.077 | -1.254 | 0.210 | -0.246 | 0.054 |
| opponent[T.M'gladbach] | -0.0646 | 0.076 | -0.852 | 0.394 | -0.213 | 0.084 |
| opponent[T.Mainz] | -0.0013 | 0.075 | -0.018 | 0.986 | -0.148 | 0.145 |
| opponent[T.Nurnberg] | 0.0927 | 0.088 | 1.056 | 0.291 | -0.079 | 0.265 |
| opponent[T.Paderborn] | 0.2886 | 0.135 | 2.130 | 0.033 | 0.023 | 0.554 |
| opponent[T.RB Leipzig] | -0.0427 | 0.118 | -0.363 | 0.717 | -0.273 | 0.188 |
| opponent[T.Schalke 04] | -0.0943 | 0.076 | -1.233 | 0.218 | -0.244 | 0.056 |
| opponent[T.St Pauli] | 0.3391 | 0.133 | 2.549 | 0.011 | 0.078 | 0.600 |
| opponent[T.Stuttgart] | 0.1592 | 0.074 | 2.144 | 0.032 | 0.014 | 0.305 |
| opponent[T.Werder Bremen] | 0.2360 | 0.071 | 3.329 | 0.001 | 0.097 | 0.375 |
| opponent[T.Wolfsburg] | 0.0355 | 0.074 | 0.479 | 0.632 | -0.110 | 0.181 |
| home | 0.2607 | 0.024 | 10.871 | 0.000 | 0.214 | 0.308 |

Out[225]:

Generalized Linear Model Regression Results

| | | | |
|------------------------|------------------|--------------------------|-----------|
| Dep. Variable: | goals | No. Observations: | 6080 |
| Model: | GLM | Df Residuals: | 6014 |
| Model Family: | Poisson | Df Model: | 65 |
| Link Function: | log | Scale: | 1.0000 |
| Method: | IRLS | Log-Likelihood: | -8475.4 |
| Date: | Fri, 07 Dec 2018 | Deviance: | 6635.6 |
| Time: | 15:23:43 | Pearson chi2: | 5.80e+03 |
| No. Iterations: | 5 | Covariance Type: | nonrobust |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|--------------------------------------|---------|---------|--------|-------|--------|--------|
| Intercept | 0.1012 | 0.120 | 0.843 | 0.399 | -0.134 | 0.337 |
| team[T.Ajaccio GFCO] | -0.0280 | 0.189 | -0.148 | 0.882 | -0.399 | 0.343 |
| team[T.Amiens] | -0.0695 | 0.189 | -0.367 | 0.714 | -0.441 | 0.302 |
| team[T.Angers] | 0.0434 | 0.130 | 0.333 | 0.739 | -0.212 | 0.299 |
| team[T.Arles] | -0.5932 | 0.237 | -2.499 | 0.012 | -1.058 | -0.128 |
| team[T.Auxerre] | 0.1500 | 0.140 | 1.069 | 0.285 | -0.125 | 0.425 |
| team[T.Bastia] | 0.0044 | 0.118 | 0.038 | 0.970 | -0.226 | 0.235 |
| team[T.Bordeaux] | 0.2135 | 0.106 | 2.013 | 0.044 | 0.006 | 0.421 |
| team[T.Brest] | -0.1777 | 0.137 | -1.297 | 0.195 | -0.446 | 0.091 |
| team[T.Caen] | 0.0377 | 0.113 | 0.333 | 0.739 | -0.184 | 0.260 |
| team[T.Dijon] | 0.1851 | 0.126 | 1.468 | 0.142 | -0.062 | 0.432 |
| team[T.Evian Thonon Gaillard] | 0.1469 | 0.119 | 1.232 | 0.218 | -0.087 | 0.381 |
| team[T.Guingamp] | 0.1129 | 0.115 | 0.978 | 0.328 | -0.113 | 0.339 |
| team[T.Lens] | -0.1384 | 0.154 | -0.901 | 0.368 | -0.440 | 0.163 |
| team[T.Lille] | 0.2600 | 0.105 | 2.466 | 0.014 | 0.053 | 0.467 |
| team[T.Lorient] | 0.1709 | 0.109 | 1.575 | 0.115 | -0.042 | 0.384 |
| team[T.Lyon] | 0.5581 | 0.102 | 5.449 | 0.000 | 0.357 | 0.759 |
| team[T.Marseille] | 0.3874 | 0.104 | 3.723 | 0.000 | 0.183 | 0.591 |
| team[T.Metz] | -0.0957 | 0.135 | -0.706 | 0.480 | -0.361 | 0.170 |
| team[T.Monaco] | 0.5276 | 0.106 | 4.988 | 0.000 | 0.320 | 0.735 |
| team[T.Montpellier] | 0.1892 | 0.106 | 1.779 | 0.075 | -0.019 | 0.398 |
| team[T.Nancy] | -0.0548 | 0.124 | -0.441 | 0.659 | -0.298 | 0.189 |
| team[T.Nantes] | -0.0904 | 0.120 | -0.755 | 0.450 | -0.325 | 0.144 |
| team[T.Nice] | 0.1865 | 0.106 | 1.753 | 0.080 | -0.022 | 0.395 |
| team[T.Paris SG] | 0.7306 | 0.101 | 7.241 | 0.000 | 0.533 | 0.928 |
| team[T.Reims] | 0.0836 | 0.121 | 0.691 | 0.490 | -0.153 | 0.321 |

| | | | | | | |
|-----------------------------------|---------|-------|--------|-------|--------|--------|
| team[T.Rennes] | 0.1374 | 0.107 | 1.284 | 0.199 | -0.072 | 0.347 |
| team[T.Sochaux] | 0.1364 | 0.120 | 1.142 | 0.254 | -0.098 | 0.371 |
| team[T.St Etienne] | 0.2190 | 0.106 | 2.067 | 0.039 | 0.011 | 0.427 |
| team[T.Strasbourg] | 0.1314 | 0.178 | 0.739 | 0.460 | -0.217 | 0.480 |
| team[T.Toulouse] | 0.0646 | 0.108 | 0.598 | 0.550 | -0.147 | 0.276 |
| team[T.Troyes] | -0.1113 | 0.136 | -0.820 | 0.412 | -0.377 | 0.155 |
| team[T.Valenciennes] | 0.0929 | 0.120 | 0.771 | 0.441 | -0.143 | 0.329 |
| opponent[T.Ajaccio GFCO] | -0.0616 | 0.151 | -0.408 | 0.683 | -0.357 | 0.234 |
| opponent[T.Amiens] | -0.3833 | 0.172 | -2.234 | 0.025 | -0.720 | -0.047 |
| opponent[T.Angers] | -0.2837 | 0.113 | -2.515 | 0.012 | -0.505 | -0.063 |
| opponent[T.Arles] | 0.1266 | 0.141 | 0.900 | 0.368 | -0.149 | 0.402 |
| opponent[T.Auxerre] | -0.2093 | 0.125 | -1.671 | 0.095 | -0.455 | 0.036 |
| opponent[T.Bastia] | -0.1353 | 0.096 | -1.412 | 0.158 | -0.323 | 0.053 |
| opponent[T.Bordeaux] | -0.3230 | 0.091 | -3.543 | 0.000 | -0.502 | -0.144 |
| opponent[T.Brest] | -0.2466 | 0.112 | -2.209 | 0.027 | -0.465 | -0.028 |
| opponent[T.Caen] | -0.0961 | 0.092 | -1.044 | 0.296 | -0.277 | 0.084 |
| opponent[T.Dijon] | 0.0590 | 0.103 | 0.571 | 0.568 | -0.143 | 0.261 |
| opponent[T.Evian Thonon Gaillard] | -0.0992 | 0.100 | -0.992 | 0.321 | -0.295 | 0.097 |
| opponent[T.Guingamp] | -0.1500 | 0.096 | -1.557 | 0.119 | -0.339 | 0.039 |
| opponent[T.Lens] | -0.0320 | 0.118 | -0.272 | 0.786 | -0.263 | 0.199 |
| opponent[T.Lille] | -0.4033 | 0.093 | -4.360 | 0.000 | -0.585 | -0.222 |
| opponent[T.Lorient] | -0.0993 | 0.090 | -1.107 | 0.268 | -0.275 | 0.077 |
| opponent[T.Lyon] | -0.3356 | 0.092 | -3.658 | 0.000 | -0.515 | -0.156 |
| opponent[T.Marseille] | -0.3835 | 0.092 | -4.154 | 0.000 | -0.564 | -0.203 |
| opponent[T.Metz] | 0.1180 | 0.102 | 1.163 | 0.245 | -0.081 | 0.317 |
| opponent[T.Monaco] | -0.4764 | 0.100 | -4.772 | 0.000 | -0.672 | -0.281 |
| opponent[T.Montpellier] | -0.2852 | 0.091 | -3.150 | 0.002 | -0.463 | -0.108 |
| opponent[T.Nancy] | -0.1700 | 0.102 | -1.674 | 0.094 | -0.369 | 0.029 |
| opponent[T.Nantes] | -0.3357 | 0.100 | -3.356 | 0.001 | -0.532 | -0.140 |
| opponent[T.Nice] | -0.2854 | 0.091 | -3.151 | 0.002 | -0.463 | -0.108 |
| opponent[T.Paris SG] | -0.6736 | 0.098 | -6.853 | 0.000 | -0.866 | -0.481 |
| opponent[T.Reims] | -0.1224 | 0.100 | -1.219 | 0.223 | -0.319 | 0.074 |
| opponent[T.Rennes] | -0.2906 | 0.091 | -3.207 | 0.001 | -0.468 | -0.113 |
| opponent[T.Sochaux] | -0.0949 | 0.100 | -0.949 | 0.343 | -0.291 | 0.101 |
| opponent[T.St Etienne] | -0.4274 | 0.093 | -4.602 | 0.000 | -0.609 | -0.245 |
| opponent[T.Strasbourg] | 0.0925 | 0.143 | 0.645 | 0.519 | -0.188 | 0.373 |
| opponent[T.Toulouse] | -0.2433 | 0.090 | -2.707 | 0.007 | -0.419 | -0.067 |

| | | | | | | |
|---------------------------------|---------|-------|--------|-------|--------|-------|
| opponent[T.Troyes] | 0.0986 | 0.102 | 0.966 | 0.334 | -0.101 | 0.299 |
| opponent[T.Valenciennes] | -0.1528 | 0.101 | -1.509 | 0.131 | -0.351 | 0.046 |
| home | 0.3028 | 0.023 | 13.118 | 0.000 | 0.258 | 0.348 |

Lets use the model to simulate matches

Manchester United vs Manchester City

Probability Man City Win (Man City Home) : 0.5213758840709862
 Probability Man U win (Man City Home) : 0.237715360715005
 Probability Draw (Man City Home) 0.2409069842947822

Probability Man U Win (Man U Home) : 0.394862588690056
 Probability Man City win (Man U Home) : 0.3464427723828131
 Probability Draw (Man U Home) 0.25869424116264067

Actual Results

1. At Etihad Stadium : <https://www.premierleague.com/match/> (<https://www.premierleague.com/match/>) (Man City wins 3-1)
2. At Old Trafford : yet to be played this year

So our model correctly predicted a Man City win

Real Madrid vs Barcelona

Probability Madrid Win (Madrid) : 0.43699669462029517
 Probability Barcelona win (Madrid) : 0.3483928560868603
 Probability Draw (Madrid) 0.21460109811900938

Probability Barcelona Win (Barcelona) : 0.6127844341417124
 Probability Madrid win (Barcelona) : 0.20240895399328018
 Probability Draw (Barcelona) 0.18475050565145482

Actual Results in 2018-2019 season :

1. Real Madrid vs Barcelona at Madrid : yet to be played this season
2. Real Madrid vs Barcelona at Barcelona : <https://www.bbc.com/sport/football/45995215> (<https://www.bbc.com/sport/football/45995215>) (Barcelona won 5-1)

Our model correctly predicted this one too