Fantasy Football

This project will focus on how to use data from the Premier League to make predctions about the coming season. This was motivated by my desire to beat my friends at Sky Fantasy Football in order to make some money.

The 3 main questions this project will look to answer are

1. How to predict the number of goals a team will score in a season
2. How to use the number of goals scored to predict the score in each match
3. How to create a metric for attack and defence for each team over a range of dates

The results of past seasons will be used throughout this project to make predictions.

```
head(results[,1:10], 3)
```

```
##    Season       Date    HomeTeam   AwayTeam FTHG FTAG FTR HTHG HTAG HTR
## 1 2005-06 2005-08-13 Aston Villa     Bolton    2    2   D    2    2   D
## 2 2005-06 2005-08-13     Everton Man United    0    2   A    0    1   A
## 3 2005-06 2005-08-13      Fulham Birmingham    0    0   D    0    0   D
```

The data object overall_goals has been used to summarise the overall number of goals scored by each team by season and average_goals is a data object which summarises the averege goals scored per game and the home:away ratios.

```
tail(overall_goals[,1:6], 3)
```

```
##      Season     Team overall_goal home_goal away_goal overall_concede
## 298 2019-20  Watford           36        22        14              64
## 299 2019-20 West Ham           49        30        19              62
## 300 2019-20   Wolves           51        27        24              40
```

average_goals - summary of goals per game by each team per season with home:away ratios

```
head(average_goals[,1:6], 3)
```

```
##    Season        Team overall_goal home_goal away_goal overall_concede
## 1 2005-06     Arsenal         1.79      2.53      1.05            0.82
## 2 2005-06 Aston Villa         1.11      1.05      1.16            1.45
## 3 2005-06  Birmingham         0.74      1.00      0.47            1.32
```

Part 1 - Predict Season Goals

Is it possible to predict the number of goals a team will score in the Premier league before the season has started?

This section will look at machine learning and regression techniques which can be used to predict the overall number of goals in the next season.

Data Preperation

In order to implement a machine learning approach the average_goals data frame has been manipulated to be shaped as a matrix which has the goals scored in the 4 previous seasons as variables and the label is the average number of goals scored in the next season.

goals_per_season is a data frame containing the average goals in the last 4 seasons by a team and will be used in order to predict the average goals scored in the current season.

```
head(goals_per_season, 3)
```

```
##             Season-4 Season-3 Season-2 Season-1 Season(y)
## Arsenal         1.71     2.03     1.95     1.92      1.47
## Bournemouth     1.18     1.45     1.18     1.47      1.05
## Chelsea         1.55     2.24     1.63     1.66      1.82
```

For now the data prepertion only includes the average goals scored in previous seasons. It is clear that more variables are required to make an accurate prediction such as league finish, manager, transfers etc.

Model

For now I do no have a model to predict the goals scored in the coming season

Future Models

The first implementation of this project will not consider any factors as the season goes on and hence will aim to predict all results in each match before the season has started. I would like to add in the future factors that consider team lineups and focus on weather the premier league scores are a stochastics process, hence, do certain teams have a pattern in form.

In future versions of the model I would like to include categorical data such as: transfers made, weather, manager, formation, lineup ...