English Premier League Exploratory Analysis

Jai D. Bansal

April 11, 2016

This report conducts exploratory analysis using English Premier League match data from the 2003/2004 season to the 2015/2016 season. The data can be obtained from: <http://www.football-data.co.uk/englandm.php>

### Summary Statistics:

* First Match Date: 2003-08-16
* Last Match Date: 2016-04-10
* Number of Seasons: 13
* Number of Matches: 4883
* Number of Teams: 38
* Number of Referees: 59

### Relegation:

At the end of each season, the bottom 3 teams in the standings are relegated to a lower division. The English Premier League (EPL) is the top English league, so teams cannot be 'promoted' up out of the league. Let's look at how many seasons the teams in our dataset have spent in the EPL.

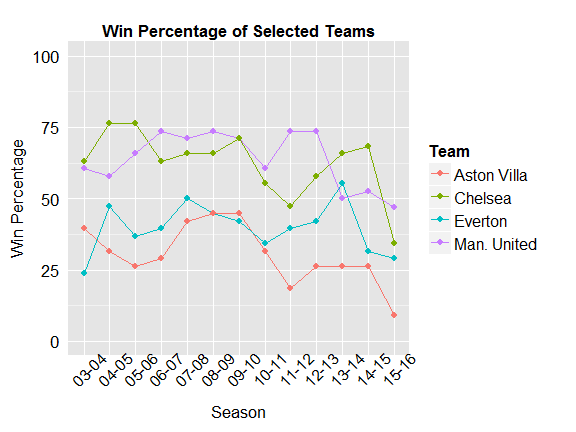
## Seasons Teams  
## 1: 1 6  
## 2: 2 2  
## 3: 3 3  
## 4: 4 4  
## 5: 5 2  
## 6: 6 3  
## 7: 7 1  
## 8: 8 2  
## 9: 9 3  
## 10: 10 2  
## 11: 11 1  
## 12: 12 1  
## 13: 13 8

Out of 38 teams in the data, 6 (15.8%) only spent 1 year in the EPL. 8 teams (21.1%) spent 13 years, the longest possible time frame in this data set. The average number of seasons spent in the EPL is 6.8. Note that large numbers of teams played 13 seasons and 1 season, with relatively low numbers of teams for the numbers in between.

This data implies that there are many teams that break into the EPL from a lower division...and are promptly relegated the following year. It also implies that there is a group of dominant teams that are rarely relegated. Looking at the clubs that played all 13 seasons in the EPL (arsenal, man united, liverpool, chelsea, everton, man city, tottenham, aston villa) reveals many famous (and wealthy) teams.

### Best of the Best and Worst of the Best:

In this section, I plot the win percentages over time of 4 teams (any more than 4 is too cluttered). Win percentage is defined as (# of wins / # of matches). I only use teams that are in the data set for every season. This is an important bias. Teams that are in the data every season are consistently above a certain quality as they do not get relegated. One could call these the best teams. I will plot the top 2 teams (best of the best) and the worst 2 teams (worst of the best).



All 4 of these teams are part of the dominant 8 that are never relegated in this data set. I find it surprising that Everton and Aston Villa (especially Aston Villa) are consistently below a win percentage of 50%. It seems that these clubs haven't done well recently, but haven't done poorly enough to be relegated. Finally, it's interesting that all 4 teams on the plot do worse, in some cases sharply, in the 2015-2016 season than the 2014-2015 season.