English Premier League: What makes a winner

Anthony Meyers, Danny Moncada, & Frankie Cancino

University of Minnesota
Driven to Discover℠

English Premier League - Contextual Overview

- # 1 Football (Soccer) League in the world - Bleacher Report
- Average per match television audience for NBC: 449,000
- Top four teams receive automatic entry into the group stages of the UEFA Champions League:
  - Team earns € 15,250,000 for qualifying
  - The 2012-2013 final at Wembley Stadium drew 360 million television viewers worldwide

Problem

- Make effective marketing decisions by understanding the factors that influence a team's final standings
- This will be vital to how teams and matches are highlighted for marketing potential
- Understanding these variables will allow effective deployment of marketing capital

Leicester City, who started the season with a **5,000 to 1** chance to win the Premier League, hoist the trophy at the end of the 2015-2016 season

Season	Champion	Runner Ups
2009–10	Chelsea †	Manchester United Arsenal Tottenham Hotspur
2010–11	Manchester United	Chelsea Arsenal Manchester City
2011–12	Manchester City	Manchester United Chelsea‡ Arsenal
2012–13	Manchester United	Manchester City Chelsea Arsenal
2013–14	Manchester City	Liverpool Chelsea Arsenal
2014–15	Chelsea	Manchester City Arsenal Manchester United
2015–16	Leicester City	Arsenal Tottenham Hotspur Manchester City
2016–17	Chelsea	Tottenham Hotspur Manchester City Liverpool
2017–18	Manchester City	Manchester United Tottenham Hotspur Liverpool

Premier League Top 4 Finishers, '09-'10 to Present

Problem

Club	Sponsor (nationality)	Value 2017-18	Industry	+/-
Man Utd	Chevrolet (US)	£47m	Cars	-
Chelsea	Yokohama (Japan)	£40m	Tyres	-
Man City	Etihad (UAE)	£35m	Airline	+£15m
Tottenham	AIA (China)	£35m	Insurance	+£19m
Arsenal	Fly Emirates (UAE)	£30m	Airline	-
Liverpool	Standard Chartered (UK)	£30m	Bank	+£5m
West Ham	Betway (Malta)	£10m	Gambling	+£4m
Everton	SportPesa (Kenya)	£9.6m	Gambling	+£4.3m
C Palace	ManBetX (Philippines)	£6.5m	Gambling	+£1.5m
Newcastle	Fun88 (China)	£6m	Gambling	n/a
Southampton	Virgin Media (UK)	£6m	Telecoms	-
Swansea	LeTou (China)	£4.5m	Gambling	+£0.5m
Leicester	King Power (Thailand)	£4m	Duty Free	+£3m
Bournemouth	M88 (Gibraltar)	£3.5m	Gambling	+£1.5m
Stoke	Bet365 (UK)	£3.2m	Gambling	-
Watford	FXPro (UK)	£3m	Finance	+£1.5m
West Brom	Palm (China)	£3m	Eco-towns	+£0.5m
Burnley	Dafabet (Philippines)	£2.5m	Gambling	+£0.5m
Brighton	American Express (US)	£1.5m	Finance	n/a
Huddersfield	Ope Sports (Malta)	£1.5m	Gambling	n/a
TOTAL		**£281.8m**	(Up from £226.5m)	

- Sponsorship deals increasing in value for top performing teams
- English Premier League also increasing advertising opportunities
- Analysis can also supplement consulting capabilities for companies looking to bid on english premier league sponsorships

Avenues of Exploration

- Identify the top teams
- Identify the top matches
- What data is out there?
  - Goals scored, yellow/ red cards received, number of saves per match, number of shots (on/ off target)
- What model do we create?
  - Multiple linear regression
- What can we conclude?
- Iterate

Data: English Premier League Results 2009 - 2018

- https://datahub.io/sports-data/english-premier-league#r
  - Dataset contains data for last 10 seasons of English Premier League including the current season (over 3,000 matches)
  - The data is refreshed weekly
- Predictors for our model (for both home and away)
  - Goals For, Home Goals Against
  - Shots, Shots on Target
  - Fouls, Fouls Against
  - Yellow/ Red Cards for, Yellow/ Red Cards against

Analyses: Different k -values

model	k	r-squared	s	p-value
All predictors	22	0.8581877	2.325087	2.20E-16
Goals and Shots	8	0.8510219	2.28347	2.20E-16
Goals	4	0.8492233	2.270806	2.20E-16
Net Goals	1	0.8442622	2.288332	2.20E-16

While the k = 4 model is more complex than the k = 1 model, with relatively similar summary statistics, the k = 4 betas provide more actionable predictors than the k = 1 model

Analyses

- Collinearity amongst many of the original predictor values
- Home & Away Goals directly linked to Shots For and Shots Against

	Home_GF	Home_ShotsFor	Home_GA	Home_ShotsAgainst
Home_GF	1.0000000	0.7131591	-0.5009005	-0.6062895
Home_ShotsFor	0.7131591	1.0000000	-0.4793260	-0.6354719
Home_GA	-0.5009005	-0.4793260	1.0000000	0.6268720
Home_ShotsAgainst	-0.6062895	-0.6354719	0.6268720	1.0000000

	Away_GF	Away_ShotsFor	Away_GA	Away_ShotsAgainst
Away_GF	1.0000000	0.7675868	-0.4525418	-0.5853232
Away_ShotsFor	0.7675868	1.0000000	-0.5485788	-0.6990074
Away_GA	-0.4525418	-0.5485788	1.0000000	0.5977395
Away_ShotsAgainst	-0.5853232	-0.6990074	0.5977395	1.0000000

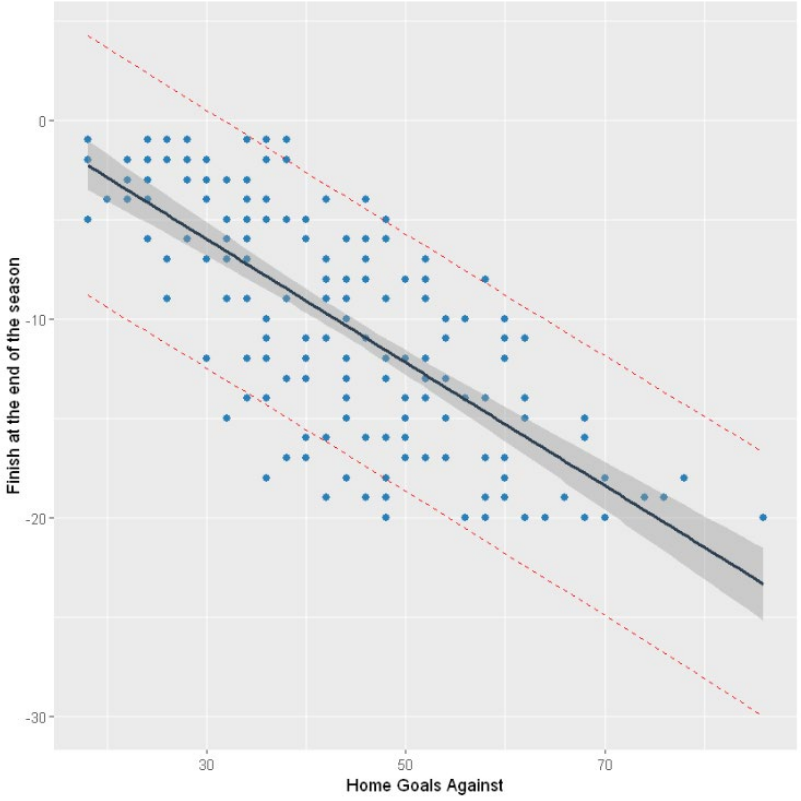
Analyses: Coefficients (k = 4)

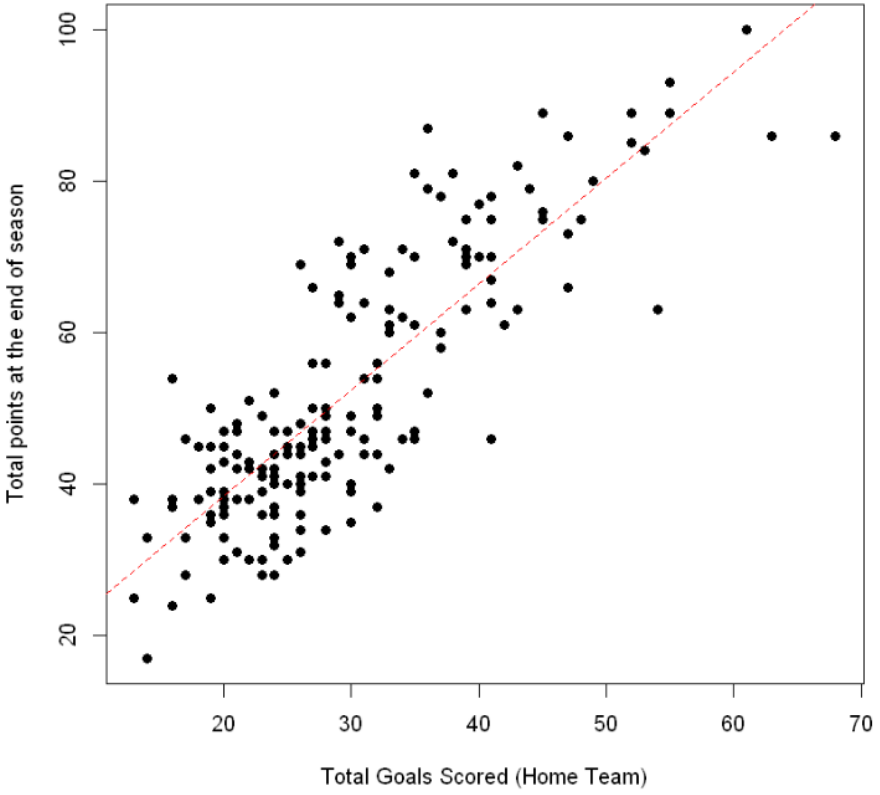
```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.87444    1.42615  -5.521 1.20e-07 ***
Home_GF      0.15763    0.02390   6.596 4.82e-10 ***
Home_GA     -0.25643    0.03165  -8.101 9.01e-14 ***
Away_GF      0.20166    0.02851   7.072 3.51e-11 ***
Away_GA     -0.20474    0.02838  -7.214 1.58e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.271 on 175 degrees of freedom
Multiple R-squared:  0.8492,    Adjusted R-squared:  0.8458
F-statistic: 246.4 on 4 and 175 DF,  p-value: < 2.2e-16
```

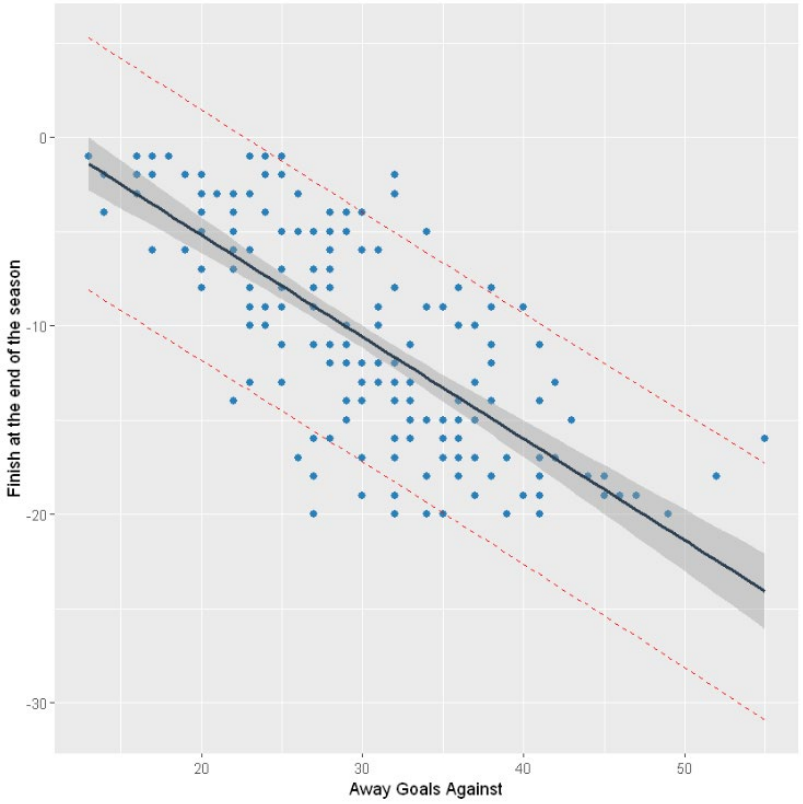
85% of variation in the final standing of a team is explained by season totals of *goals for* and *goals against*, split by home and away matches, respectively

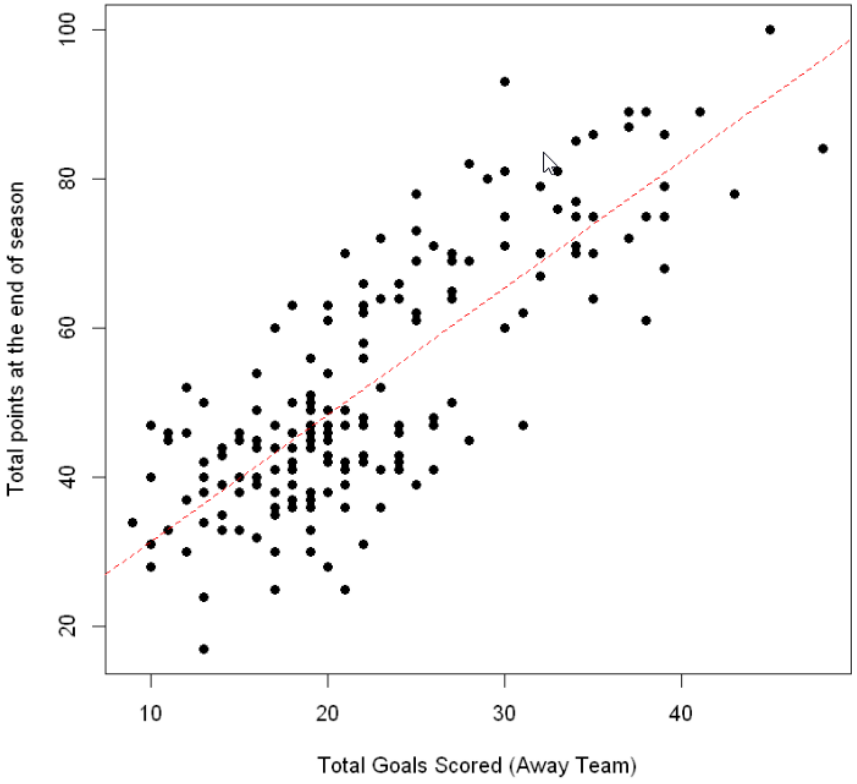
Analyses: Home Form





Analyses: Away Form





Analyses: Coefficients (k = 4)

				90% Confidence Interval	
Predictor	**Description**	**β**	**Estimate**	**5%**	**95%**
Home_GF	Goals for at home stadium	β1	0.15763	0.11811	0.19714
Home_GA	Goals against at home stadium	β2	-0.25643	-0.30877	-0.20408
Away_GF	Goals for at away stadiums	β3	0.20166	0.15450	0.24881
Away_GA	Goals against at away stadiums	β4	-0.20474	-0.25167	-0.15781

Confidence intervals define the range in which the true population effect of the predictors will lie

Conclusion

- The estimates from the linear model provide predicted value changes in the final standing of a team based on goals for and goals against as the home and away team, respectively
- These predictors can be utilized to efficiently deploy marketing capital to the team segments (cities, fans, etc.) to maximize earned revenue or to maximize broadcasting capital to potentially high profile matches

Outcomes

- Based on the k=4 model of goals, from our sampling:
  - Scoring a goal as the away team is 27.9% more valuable than a home goal to a higher finish
  - Conceding a goal at home is 25.2% more damaging than conceding a goal away
- Consider heavier marketing towards teams that score at high rates as the *visiting team* and that do not concede goals *at home*

Follow-up

- With further research and data, predictors of the *goals for* and *goals against* predictors could supplement the analysis and allow for prediction and forecasting capabilities
- Further evaluation of market research can compliment analytical approaches for counter cases that fall out of the scope of the analysis results

Questions?