RSM384: Research Project

Vivaan Bhaskar

TABLE OF CONTENTS

01

RESEARCH QUESTION

RESEARCH QUESTION:

Does the **presence of crowds** influence **home game advantage** in the **English Premier League**?

OVERVIEW OF
HOME GAME ADVANTAGE

Home game advantage is a sports phenomenon which describes the benefits the home team has over the away team

Home game advantage increases the chances of a home team victory through increased goals and referee bias

Soccer has been found to have the largest home game advantage across all leagues

02

DATA & METHODOLOGY

THE DATA

EPL Data	Attendance Data

• **EPL Match by Match performance variable**s over 2009/10 - 2020/21 season
  ○ **Existing variables include:** Home Team Indicator, goals, conceded goals, yellow and red cards, fouls.
  ○ **Created new variables:** Conversion Rate, Manipulated Cards Variable, Proxy for Referee Bias (ie. cards to fouls ratio)
• ***Source:*** https://www.football-data.co.uk/englandm.php
	• **Crowd attendance per EPL team** from 2009/10 - 2020/21 season
	  ○ **Averaged out** attendance per team per season.
	  ○ **Created a new variable**: Category of Team (dependent on appearances in last 12 PL seasons)
	  ○ Allows us to analyze the **impact of crowds** on HA and other factors
	• ***Source:*** https://www.kaggle.com/datasets/joovasco/premier-league-attendance-from-1949-to-2019 https://www.worldfootball.net/attendance/eng-premier-league-2019-2020/1/

THE METHODOLOGY

We answered our research question through a three facet approach:

EXPLORATORY

- Analyze only EPL data for performance variables.
- **Aim:** Find evidence of home advantage in EPL.
- **Method:** Study data over past 12 seasons to look for patterns.

REGRESSION

- Contribution to **Winning at Home**.
- **Aim:** Understand the impact of performance variables on home win %.
- **Method:** Multiple Regression Model
- Control 1: category of team, consider only home data.
- Control 2: home and away, consider all data.

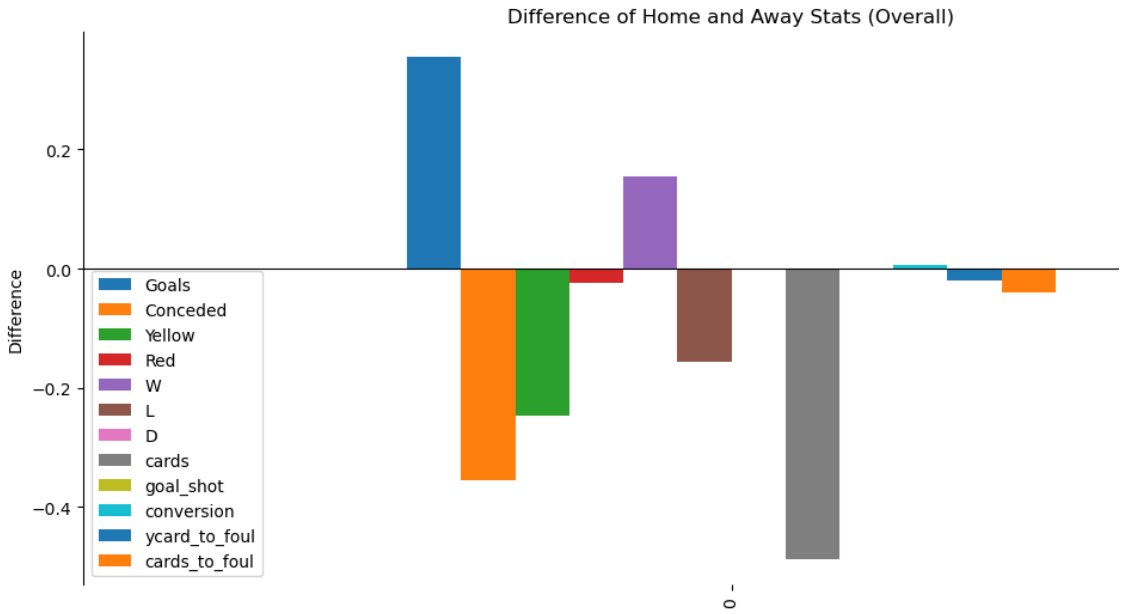
EFFECT OF CROWDS

- Build a model to understand the effect of crowds on Winning at Home.
- **Aim:** Understand the impact of crowd presence on HA.
- **Method:** Differences in regression model pre-covid vs when games were played behind closed doors.

03

VISUALIZATIONS

OVERALL ANALYSIS OF HA



MORE GOALS

Approximately 0.4 more goals on average.

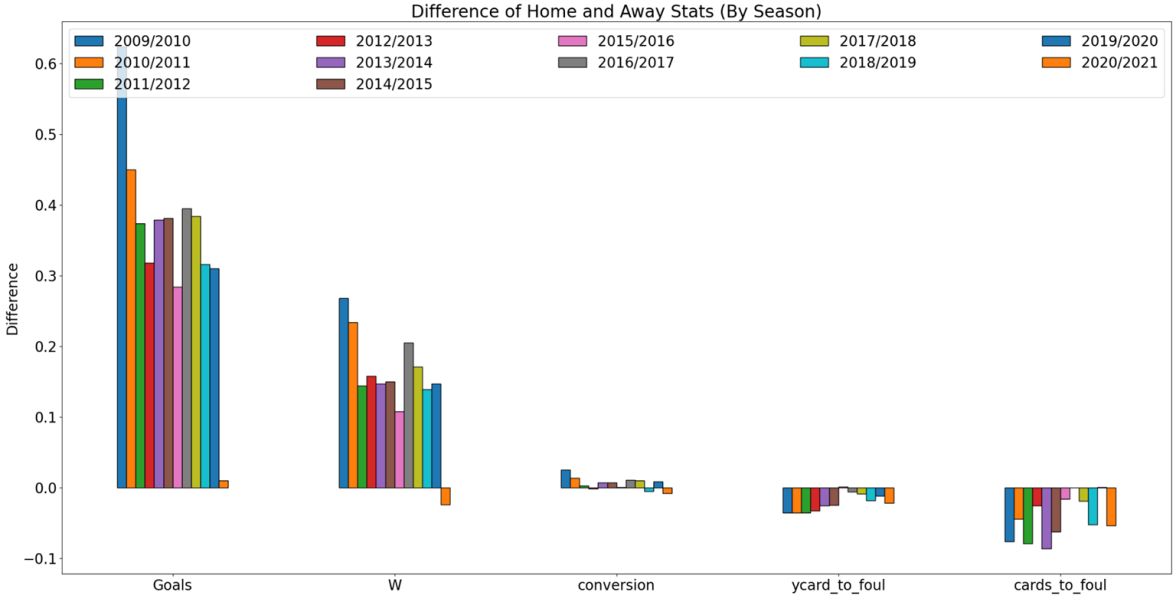
LESS CARDS

Less yellow and red cards and lower card to foul rate.

HIGHER WIN RATE

Approximately 15% higher win rate at home.

IMPACT OF HA BY SEASON



Focusing on the COVID season (2020/21):

LESS GOALS

Drastically lower goals compared to previous seasons.

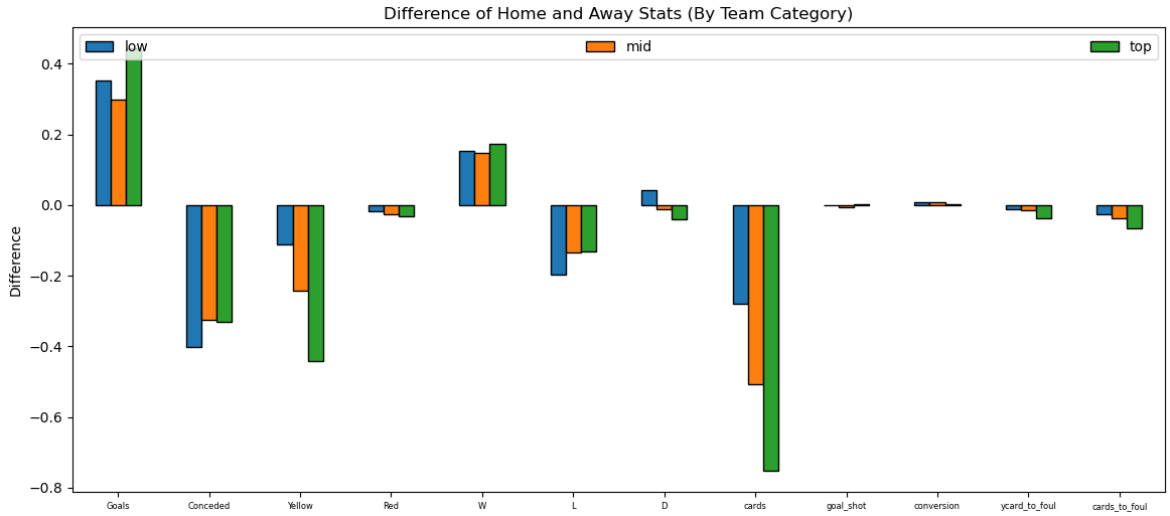
LESS CONVERSIONS

Negative conversion rate of goals.

LOWER WIN RATE

First negative home win rate in past 10 years.

IMPACT OF HA BY TEAM



TOP TEAM HA

Top teams score and win more at home.

OFFENSIVE IMPACT

Top teams get more than twice less yellow cards at home.

UNDERDOG EFFECT

Lower teams draw more away than at home.

04

REGRESSION MODEL

MULTIPLE REGRESSION #1

	Model 4	Model 5	Model 6
const	0.23***	0.34***	0.21***
	(0.02)	(0.02)	(0.03)
Att_10000		0.03***	0.01*
		(0.01)	(0.00)
Goals	0.21***		0.20***
	(0.02)		(0.02)
Conceded	-0.19***		-0.18***
	(0.02)		(0.02)
conversion	0.43*		0.48**
	(0.23)		(0.23)
cards_to_foul		-0.36***	0.03
		(0.13)	(0.07)
top	0.09***	0.26***	0.08***
	(0.01)	(0.02)	(0.01)
mid	0.07***	0.04***	0.06***
	(0.01)	(0.01)	(0.01)
low	0.07***	0.04***	0.06***
	(0.01)	(0.02)	(0.01)
R-squared	0.88	0.51	0.88
R-squared Adj.	0.88	0.50	0.88
R-squared	0.88	0.51	0.88
No. observations	233	233	233

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

INDEPENDENT VARIABLE

Home team winning percentage

MODEL 5

Home team being a "Top" team results in win percentage increasing by 26 PP

R explains 51% of the variation

Win percentage increases by 3 PP per every 10,000 people

MODEL 6

Home team being a "Top" team results in win percentage increasing by 8 PP

Goal scored by home team increases winning percentage by 20 PP

MULTIPLE REGRESSION #2

	Model 4	Model 5	Model 6
const	0.27***	0.27***	0.27***
	(0.02)	(0.03)	(0.02)
Att_10000		0.04***	0.00
		(0.00)	(0.00)
Goals	0.24***		0.24***
	(0.01)		(0.01)
Conceded	-0.17***		-0.17***
	(0.01)		(0.01)
conversion	0.15		
	(0.16)		
cards_to_foul		-0.39***	
		(0.11)	
HomeAway	0.01	0.14***	0.01
	(0.01)	(0.02)	(0.01)
R-squared	0.88	0.32	0.88
R-squared Adj.	0.88	0.31	0.88
R-squared	0.88	0.32	0.88
No. observations	466	466	466

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

INDEPENDENT VARIABLE

Home team winning percentage

HOMEAWAY VARIABLE

Winning percentage increases by 14 PP at a 1 percent significance level

05

IMPACT OF CROWDS

MULTIPLE REGRESSION #3

	Model 1(pre)	Model 1(covid)	Model 2(pre)	Model 2(covid)	Model 3(pre)	Model 3(covid)
const	0.23***	0.17**	0.34***	0.25***	0.21***	0.19*
	(0.02)	(0.08)	(0.03)	(0.08)	(0.03)	(0.09)
Att_10000			0.03***	4.12**	0.01*	0.29
			(0.01)	(1.55)	(0.00)	(1.24)
Goals	0.21***	0.20***			0.21***	0.20***
	(0.02)	(0.05)			(0.02)	(0.05)
Conceded	-0.19***	-0.13**			-0.18***	-0.11
	(0.02)	(0.06)			(0.02)	(0.07)
conversion	0.42	0.52			0.43	0.39
	(0.27)	(0.41)			(0.27)	(0.49)
cards_to_foul			-0.34**	-0.77	0.04	-0.27
			(0.14)	(0.46)	(0.07)	(0.31)
top	0.09***	0.07*	0.26***	0.15**	0.08***	0.07
	(0.01)	(0.04)	(0.02)	(0.05)	(0.01)	(0.04)
mid	0.07***	0.05	0.04***	0.04	0.06***	0.05
	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)	(0.04)
low	0.07***	0.05*	0.04**	0.05	0.06***	0.06*
	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.03)
R-squared	0.88	0.89	0.51	0.65	0.88	0.90
R-squared Adj.	0.88	0.86	0.50	0.56	0.88	0.84
R-squared	0.88	0.89	0.51	0.65	0.88	0.90
No. observations	213	20	213	20	213	20

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

HIGHER WIN RATE

Model 1: *9%*** higher win rate pre-covid, 7%* higher win rate during covid*

Model 2: *26%*** higher win rate pre-covid, 15%** higher win rate during covid*

Model 3: *8%*** higher win rate pre-covid*

06

CONCLUSION

RESEARCH QUESTION:

Does the **presence of crowds** influence **home game advantage** in the **English Premier League**?

YES!

HOME ADVANTAGE IS PRESENT

MORE GOALS

0.4 more goals on average for top teams

HIGHER WIN RATE

Approximately 15% higher for top teams

LESS CARDS

Receive twice less yellow cards for top teams

UNDERDOG EFFECT

Lower teams draw more away than at home

IMPACT OF CROWDS

Higher Winning Rate for top teams with 26% higher rate pre-COVID and 15% higher during COVID

Limitation to this finding as the number of observations during the COVID season is small

LIMITATIONS OF OUR APPROACH

CONFOUNDING VARIABLES

Home Team Performance during Covid could have reduced due to other reasons - bad transfer market, drop in coaching level, drop in player performance, etc.

AVERAGE ATTENDANCE

It would be beneficial to have match by match attendance data, rather than average attendance, which could further our analysis.

ENDOGENEITY BIAS

Regression Models suffer from EB (ie. independent variable is correlated with error term)

THANK YOU