

Exploring statistics of Premier League 2018/19*

Dai Moroi

April 6th 2022

Abstract

This paper analyzes the basic statistics of England professional soccer league, Premier League, to figure out what makes the team strong and also examines statistical trends that the league has. The result indicates that strong teams are superior in terms of the percentage of goal scored out of total number of shots, while average possession doesn't really matter to game results.

Keywords: football, soccer, Premier League

1 Introduction

What makes the team strong? What kind of characteristics does a strong team have? To answer those questions with convincing stories, statistics can play an essential role. Since the Internet enabled us to accumulate data easily, use of data has been featured in the world of sports, including soccer. This paper takes the example of 2018-19 season of England professional soccer league, Premier League, and examines its basic statistics, such as goals, possession rate, the number of shots, etc., from various aspects.

This paper uses data from a website, [FootyStats] (<https://footystats.org/>), and the analysis for this paper uses the R statistical programming language (R Core Team 2021). First, I explore the overview of the statistics to understand the characteristics of the league. Then, I will examine which kind of data is most positively correlated with the league position to figure out the feature of strong teams by plotting data and calculating the correlation coefficient between league position and some particular data.

The remainder of this paper is structured as follows: Section 2 discusses the data using graphs and tables. This section also explains the methodology that the data was collected and bias they have. Section 3 discusses the model, Section 4 presents the results, and finally Section 5 discusses the findings and weaknesses to be overcome.

2 Data

3 Model

4 Results

Discussion

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

*Code and data are available at: github.com/moroidai/Final-Paper.