Soccer Data Analysis: Report

by Joseph E. Rehfus

Table of Contents

Motivation

The English Premier League (EPL) is the most-watched sports league in the world. Each season, it features 20 professional soccer clubs vying for the prestigious league title and for qualification to compete with the best European clubs in continent-wide tournaments. In addition to the battle to secure glory and riches amongst the top teams, there is a perilous battle for survival between the worst clubs, the bottom three of which will be relegated to a lower division at the end of the season, losing out on revenue and prestige. The EPL season is full of drama from start to finish, which is likely why its storylines attract such diverse and international viewership. As a fan, I wanted to use the match results data to answer a few exploratory questions about the EPL.

Scoring Goals

Data requirement
- final scores from EPL matches
- the time at which each goal was scored in EPL matches

Data source(s)
- www.worldfootball.net

Pertinent script(s)
- plot_league_goals.py

How many goals are typically scored during an EPL match?
What does the distribution of total goals scored look like?
Text.

What is the typical margin of victory when one team beats another?
Text.

When are goals scored during a match?
Text.

Determining League Position

Data requirement: the result of each league match in a season
Data source: www.worldfootball.net
Pertinent script(s): plot_league_points.py

Data requirement
- final scores from EPL matches

Data source(s)
- www.worldfootball.net

Pertinent script(s)
- plot_league_points.py

When are league points won throughout a season?
Text.

How does goal differential develop throughout a season?
Text.

How do total goals scored develop throughout a season?
Text.

How does league position change throughout a season?
Text.

Conclusions