

Origin of data:

My data is from [1] an online website that stores historical data on football matches for different leagues. The league that I chose was the English Premier League. I aggregated data from multiple seasons, beginning from 2009/2010 until 2018/2019, resulting in data from nine seasons in the Premier League. The full aggregated .csv file is included with my submission under the title 'FullBettingDF.csv'.

Main findings:

The main findings I found were the effects that different variables such as the number of shots the home or away team took impacted the full time result. What was really interesting to see was how much an advantage was given to the team playing at home, as they could have been completely dominated in terms of shots, however still had a chance of winning, simply by having the advantage of being at home and having the support of the home fans.

When performing the explanatory data analysis, I wanted to include plots which were simple, however showed a lot of information. I had done a couple facet grids in the exploratory analysis but chose not to select them as they took up too much space, and it was difficult to compare the grids together.

Sources:

[1] <https://www.football-data.co.uk/englandm.php>