Final Project Analysis

Impact of Fans in Stadiums:

Results from the top 5 European leagues going back 5 seasons were analysed to determine what, if any effect the absence of fans during COVID-19 lockdowns had on the frequency of home teams winning. We found that, on average home teams won 44.3% of matches in which fans were present and only 40.2% during the lockdown period.

Average Goals per Match by League:

The average number of goals scored per match was calculated for each of the top 5 European leagues as well as the Champions League over the last 5 seasons. The results showed that the median total for 5 out of 6 leagues is 2.5 goals per game. The Spanish league is the exception at only 2.0 goals per game. The French league was found to be the most variable is terms of results. Outliers were calculated both manually and using a boxplot. It was determined that games with 7 or more total goals were considered outliers. 69 out of our 9656-match sample were classified as outliers.

Squad Wages Analysis:

A 5-year average squad wage was calculated for each team that participated in any of the last 5 seasons in the top 5 European leagues. We then used this along with the average point tally of each team over the last 5 seasons to find the ratio of points earned per million dollars spent on player wages. Interestingly, only teams from France feature in the top 10 of this metric, while English teams make up 9 of the bottom 10 on this list. A scatter plot with linear regression was then generated with points as the x-axis and wages as the y-axis. The result shows a weak correlation between wages and points earned (r^2 = 0.5).

Actual Goals vs Expected Goals Analysis:

In addition to wages, scatter plots with linear regression were also generated for ball possession percentage, goals scored, expected goals scored, goals conceded, and expected goals conceded. Of these 5 variables, possession was least correlated with points earned (r^2 = 0.62) and goals scored was most correlated (r^ = 0.86). The r^2 values of the expected goal metrics versus the actual goals were compared to determine how reliable the expected stats are when it comes to predicting future outcomes. A difference of 0.08 in r^2 values tells us that the expected metrics are generally accurate but still imperfect.

Average Ages/Wages Across Leagues:

When grouping the summary dataframe by league we were able to generate 2 bar graphs to show the average player ages and wages across each of the top 5 European leagues. We found no significant difference in average age across any of the leagues and that the EPL has far and away the highest average wages.

Goals Scored & Wages Year over Year:

Finally, the summary data was grouped by season so that we could uncover any year-to-year trends in player wages and total goals scored in each league. We found no discernable trend in total goals scored, with many of the leagues experiencing wild fluctuations season to season. Wages on the other hand continue to rise at a steady rate in the EPL, while seemingly plateauing in every other league.

Implications:

This analysis takes a step towards quantifying the effect of ‘home field advantage.’ Based on our linear regression analysis we can say that teams who implement a more offensive style of play are more likely to succeed relative to expectations compared to those who opt to play more conservatively. This makes sense when you consider that the margin of error becomes increasingly small when a team plays in lower scoring games on average. Across all the leagues there seems to be little correlation between spending and team performance, however this relationship could be obscured by the fact that the EPL continues to separate themselves from the rest of Europe to the point where the average bottom table English team still has a much more expensive squad than top tier teams from other countries. This widening gap could prove catastrophic for the rest of the top European leagues as top talent will increasingly look to make the move to England due to financial incentives.