

Milestone 2 Writeup

Comparing Home and Away Teams: Kyle Cleveland

One way in which we explored our dataset was comparing the statistics of home teams to the statistics of away teams for a given match. While exploring the difference in statistics between home and away teams, there were some results that were expected but there were also some interesting insights that were discovered. Home field advantage was certainly expected which was shown in the data. Home teams won 47.63% of matches while away teams only won 33.68% of matches. This confirmed our team's hypothesis that the location of a match (whether a team is home or away) has a significant impact on the outcome of the match. One surprising insight from exploring home and away team data was that the percentage of shots on target was roughly the same for home and away teams. Away teams actually had almost 2% more of their shots on target than home teams did but that difference is pretty negligible. The difference in shot data between home and away teams actually came from the total number of shots attempted by a team in a match. On average, home teams attempted 2.99 more shots than away teams in a given match, we were also able to see the difference in distributions of shots taken by creating boxplots. Another surprising insight was the number of fouls and cards earned by home and away teams is pretty similar. While away teams on average were called for slightly more fouls (.15 more) and earned slightly more cards (.19 more) the distributions of home fouls and cards compared to away fouls and cards were very similar, so this difference is relatively negligible.

The exploration of home and away team data should greatly influence the direction of our project. We have confirmed the presence of home field advantage and know the likelihood of a home team winning, drawing, or losing a match. We've also learned that home and away teams have relatively the same percentage of shots on target and commit roughly the same amount of fouls and earn the same amount of cards. But there is a significant difference between the average number of shots home and away teams take. These correlations and insights will be useful when we build our model that predicts the result of a given match and could be useful even for predicting match statistics.

Match Results:

Another way we explored our dataset was by looking at the impact of various stats on the scores of the match. One of the things we looked at was the impact of yellow cards and red cards on the total match score. Initially, we thought that the greater the number of yellow and red cards, the higher the total match score would be. Our thought process for this was that there would be more goal-scoring opportunities with a player imbalance

from a red card, or more goal opportunities with a player playing more cautiously to avoid getting a red card (if they had a yellow card). However, this hypothesis was unfounded when we looked at the actual data. For yellow cards, there was an even distribution to which yellow cards per game were not an accurate predictor of total goals per game. For red cards, there was more of a correlation, however, this was also a relatively even distribution and didn't have enough matches with red cards to really find any sort of patterns within the data. Perhaps the most interesting statistic we found was finishing accuracy. We calculated this by taking the total shots from both the home and away teams as well as the total goals per game. Dividing the latter by the former gave us a finishing percentage that allowed us to determine how good both teams were at capitalizing on goal scoring opportunities. In the future, the finishing percentage could be a very valuable statistic to look at in our project. The finishing percentage of a team could help determine their likelihood of winning a match when given goal scoring opportunities.