# Can data analysis provide an advantage in football sports gambling?

Scott Thomas, Ken Saville, Kirk Novak

**Introduction**:

Sports gamblers are always looking for an edge. However, oftentimes this perceived edge is based on feelings of hunches rather than on solid information. For example, a common misconception about sports betting is known as the gamblers fallacy, which can be summarized as: "a team on a losing streak is guaranteed to win their next game". That is, they are "due ".

In reality, sports betting is carefully designed by the bookmakers, or "the house" to come out ahead. The bookmakers calculate score spreads and over/under lines in such a way as to encourage bettors to evenly place bets on both sides of the equation. Because the bookmakers charge a small fee (or juice) for each bet, if they tailor the odds to promote a 50/50 success rate, they will always come out ahead.

So when placing bets, particularly in sports, where certain available information could affect the outcome, as opposed to a dice or card game where the odds are essentially a fixed entity, it is may be possible to research the teams and conditions involved to slightly increase the odds of winning.

This idea is illustrated by the following quotes from <http://www.lootmeister.com/betting/quotes.php>.

“Remember this: The house doesn’t beat the player. It just gives him an opportunity to beat himself.”--old professional gambler Nicholas Dandolos

“Bet with your head, not with your heart.”--proverb

As well as this advice from football gambling expert Tim Johnson:

"The best defense against making bad decisions is always the time you should take to do your research before a play. With so much information available, there’s no excuse for betting on a team based on your hunches or feelings." - Tim Johnson / Author

As an exercise in data analysis, we chose to analyze some football score and betting data to investigate some of the information that might be taken into consideration when making football bets. The data for this analysis was obtained from https://www.kaggle.com/tobycrabtree/nfl-scores-and-betting-data. Below is the desription of this data from Kaggle.

National Football League (NFL) game results since 1966 with betting odds information since 1979. Dataset was created from a variety of sources including games and scores from a variety of public websites such as ESPN, NFL.com, and Pro Football Reference. Weather information is from NOAA data with NFLweather.com a good cross reference. Betting data was used from http://www.repole.com/sun4cast/data.html for 1978-2013 seasons. Pro-football-reference.com data was then cross referenced for betting lines and odds as well as weather data. From 2013 on betting data reflects lines available at sportsline.com.

Our goal was to investigate the performance of teams relative to the betting spread and over/under line as well as to investigate the potential effects of weather on overall scoring. Some of the questions we addressed were: Historically speaking are certain teams better against the spread than others?

Is there a home field advantage? Tat is, is there an overall scoring advantage for the home team?

Does weather affect score? Are scores higher at moderate temperatures and/or lower windspeed? If so is this affect more pronounced for the home team over the away team?t

Is over or under a safer bet? That is, do the total scores of games, historically, tend to be higher or lower than the over/under line set by bookmakers?

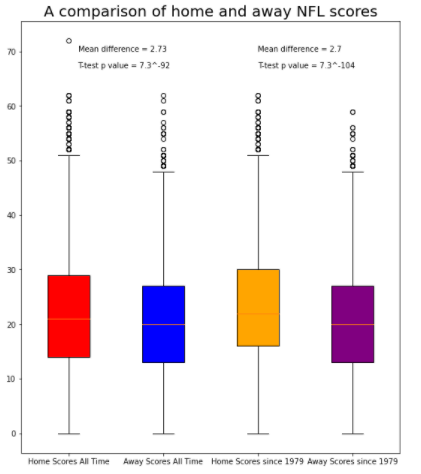
When should I take the points? Or, in other words, are there certain teams that tend to 'cover the spread'? Is there an effect of home field advantage on this metric?

Analysis:

*Is there a home field advantage in scoring?*

We compared the mean scores for home and away teams for all games from 1966 through the 2020 season (N = 12,934) as well as the subset of these games from 1979 to present, as these games also had betting data (spreads and over/under lines) available for subsequent analyses. We found a mean difference in home vs away scores of 2.7 for each comparison (Figure 1). An independent T-test comparing these means found that this was a statistically significant difference.

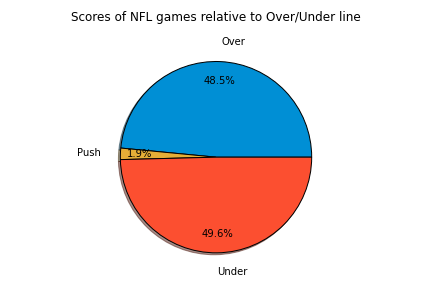
Figure 1.



## Is over or under a safer bet?

For each game since 1979, we calculated the combined score for both teams and compared this total score to the over under line for that game. The percentage of games that were "over", "under" or a "push" are shown in Figure 2. There was no apparent difference between these outcomes (48.5 % were over, 49.6% were under, and 1.9% were a push). So, the bookmakers very successfully calculated the over underlines to result in about a 50/50 split. It would be interesting to compare these numbers to the number of bettors that chose these different outcomes.

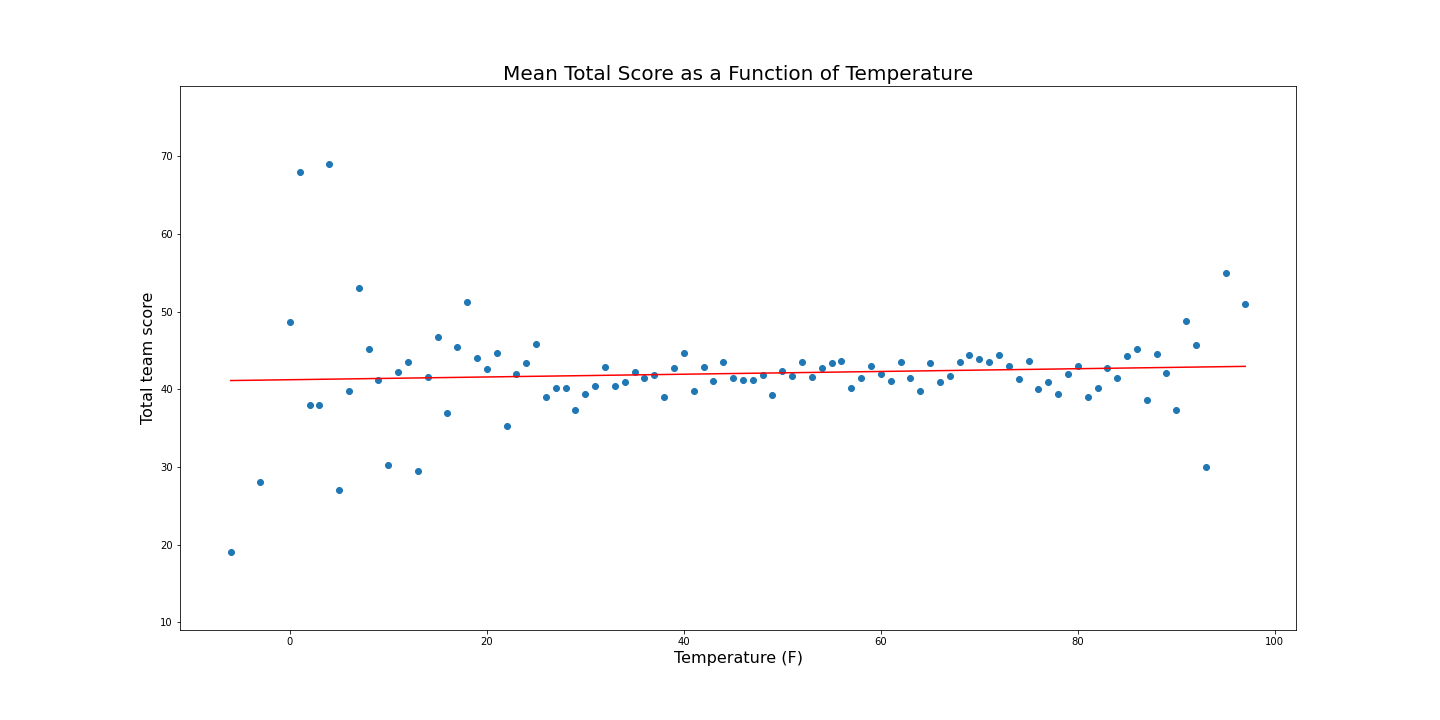
Figure 2.

**

*Does temperature affect score?*

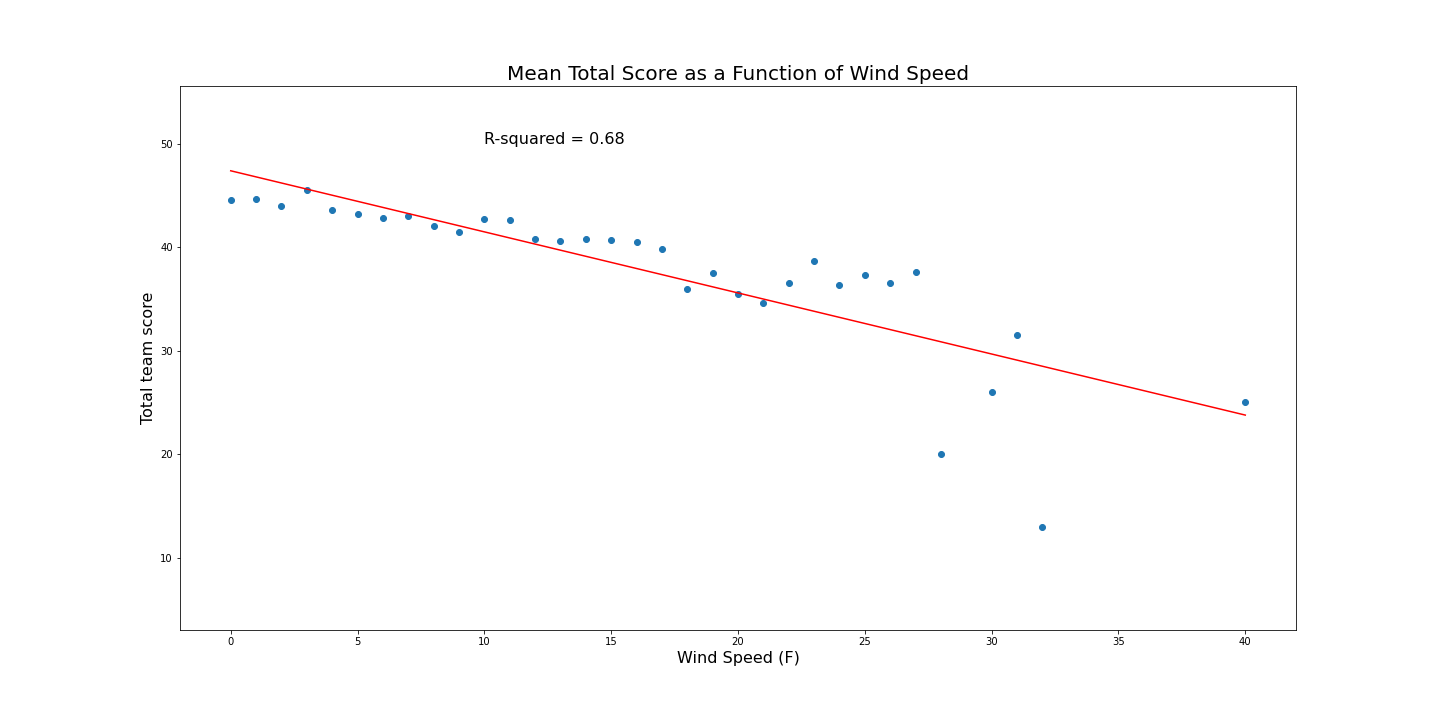
We next analyzed the correlation between game time temperature and total score to see if knowing the temperature in advance might provide some insight into whether to bet the over or under. As shown in Figure 3., there was no correlation between temperature and total score. There was more variability at lower and higher temperatures. While this may in part be due to fewer games at these extremes, it might also point to some opportunity to exploit this information. However it does appear that lower and higher scores are also evenly distributed at these extremes, just that the magnitudes of the differences are greater. We also did these analyses for home and away team scores and also found no correlation (not shown).

Figure 3.



*Does wind speed affect total scores?*

Similar to the above analysis, we looked at total game scores as a function of wind speed (Figure 4). In this case we did observe a statistically significant negative correlation between score and wind speed. That is, as might be expected, total scores tended to be lower in higher wind conditions. While this might suggest to the casual observer that an 'under' bet would have better odds of winning, it would be prudent to consider that the bookmakers also have access to this information. It would be interesting to see if the over/under lines were historically set at lower numbers in windy games.

Figure 4. 

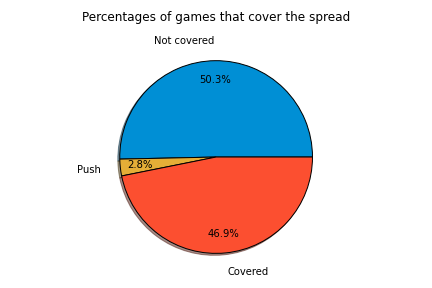
## Should I take the points?

The major betting parameter when betting on football games is the "spread". That is, the bookmakers designate a favorite team and assign a number of points (the spread) by which that team is favored. Bettors picking the favorite must 'give' points equal to the spread, and bettors picking the underdog are said to 'take' the points. If the favorite wins the game by more points than the spread, they are said to have 'covered' the spread. Is there a way to determine which of these strategies (giving or taking points) has better odds in a given game? One way to approach this is to analyze each team to see if they tend to 'cover' the spread or not.

*overall percentage of teams covering the spread.*

We first analyzed the percentage of teams that have covered the spread since 1979 (Figure 5). As was seen with the over/under percentages, the bookmakers tend to get this right, with about a 50/50 split for teams covering the spread or not (46.9% cover, 50.3% not covered, 2.8% push).

Figure 5.



*Historically speaking are certain teams better against the spread than others?*

We analyzed the tendency of each NFL team to cover the spread (Figure 6). Similar to the above analysis, the mean percentage of spreads covered was about 47% (horizontal line in Figure 6). While a few teams (notably Green Bay and San Francisco) slightly exceeded the average, and a few (Tampa Bay, Tennessee and Washington) were a bit below the average, there was very little overall variation from the mean.

Figure 6.

*A picture containing icon

Description automatically generated*

*Does playing at home or away affect teams' performance against the spread?*

We next broke down the spread analysis to take into account whether or not teams played at home or away (Figures 7 and 8). As above, most teams were about 50/50 covering away or at home with a few minor exceptions. Buffalo, Green Bay and Pittsburgh had a seemingly significant tendency to cover at home, while Indianapolis, New Orleans and San Francisco tended to cover the spread on the road.

Figure 7

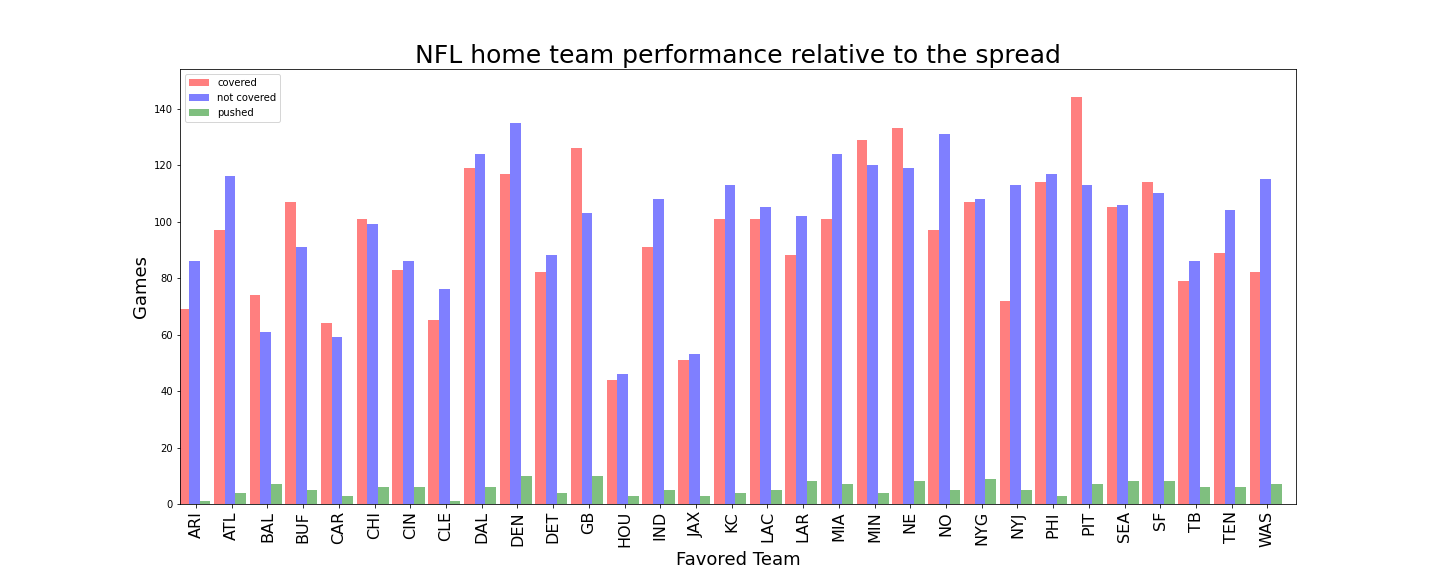
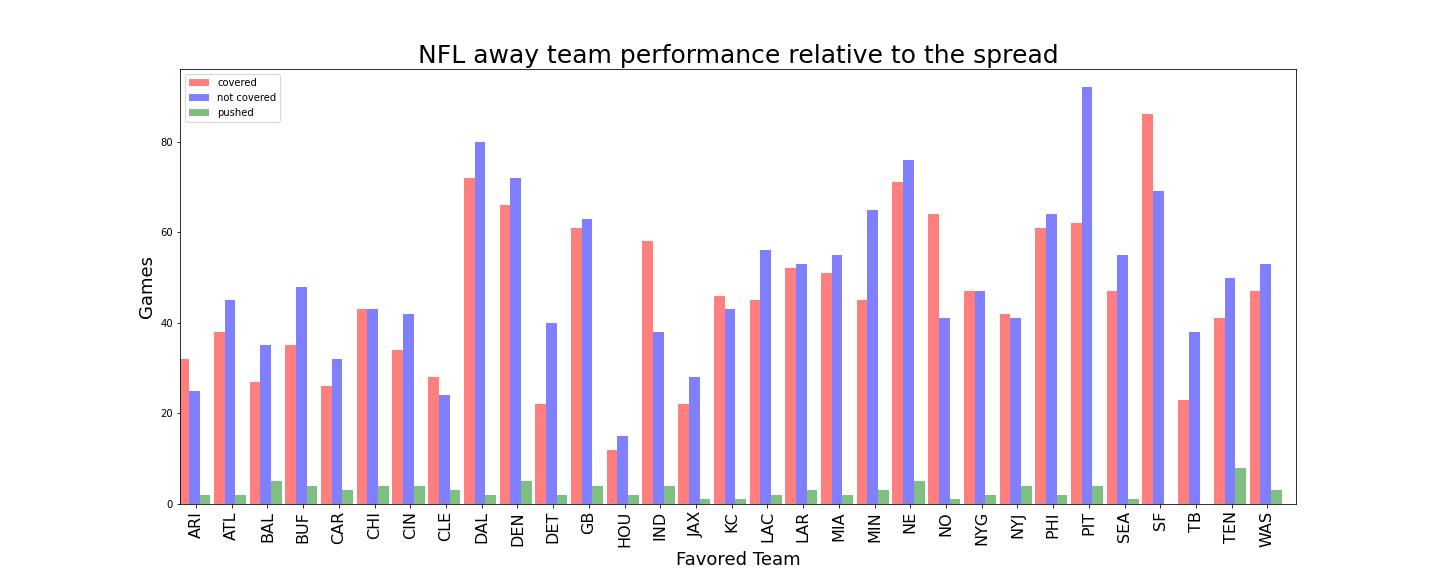


Figure 8.



## Conclusions and Future Directions

While this was a good exercise for accessing, analyzing and presenting some data related to football game scores and betting data, no major insights promising to increase our chances of betting success emerged from the analysis. Additional data to take into consideration for future analyses could include Starting quarterback status, time of game (particularly games involving teams from the east and west coast playing in the other time zone), teams coming off bye weeks, and injuries of star players. However, it is likely that bookmakers have taken these and thousands of other factors into consideration and it is best to approach sports betting as a form of entertainment and not as a potential money-making venture. Even if we were to identify some potential advantage it would be best to keep the following adage in mind: If it looks too good to be true, it probably is.