# NFL Pick Model

Background: NFL pick em pools require participants to pick games against the spread. In theory, in a perfect world with perfect lines, there would be approximately 50% of participants picking each team against the spread. In reality, one team will be chosen by 60%, 70%, or even 80%+ of participants.

Thesis: "Crowded" picks often yield unexpectedly poor results - my anecdotal experience has shown that 80% is the approximate tipping point. Is there a linear relationship or threshold at which it is beneficial to go *against* the "wisdom" of crowds?

Initial analysis has not found correlation between pre-pick metrics (team avg point/yard/turnover differential over last 3 games, performance against the spread over the last 3 games). I am continuing to try and improve factors.

However, the single data point of pick percentage does seem to support the hypothesis.

This graph plots x = Pick Pct (rounded to nearest 0.05) against adjusted outcome (win margin relative to spread, converted to # of possessions (ceiling of (margin / 7)).

Boxplot grouped by ats_pct_adj
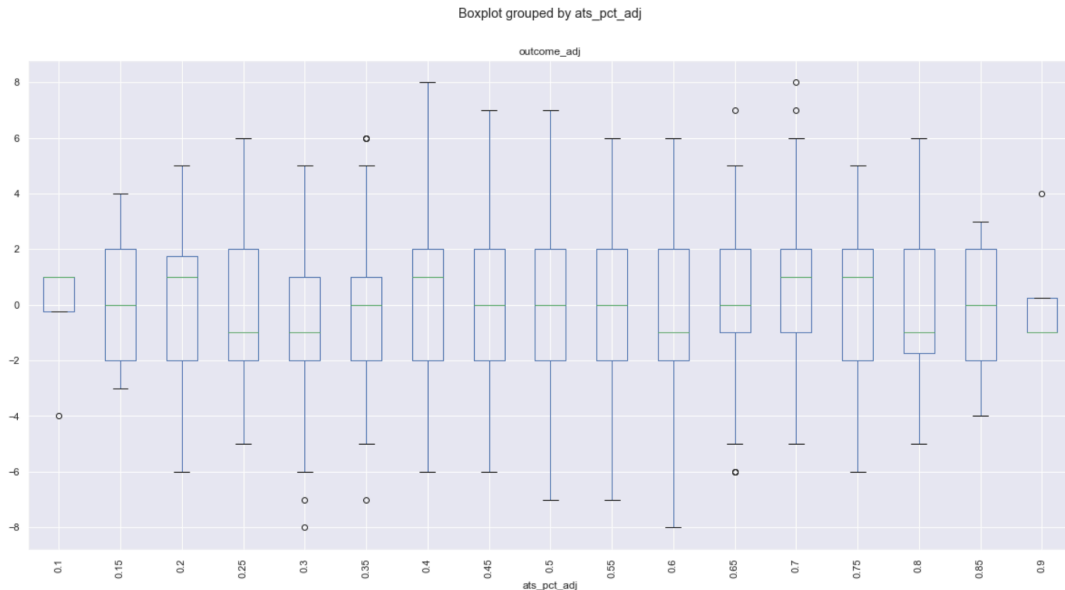
Additionally, some of the individual Boolean features seem to have a relationship that requires further investigation.

This graph plots x = 0 or 1, with 1 meaning that the team has failed to cover 3 straight games.
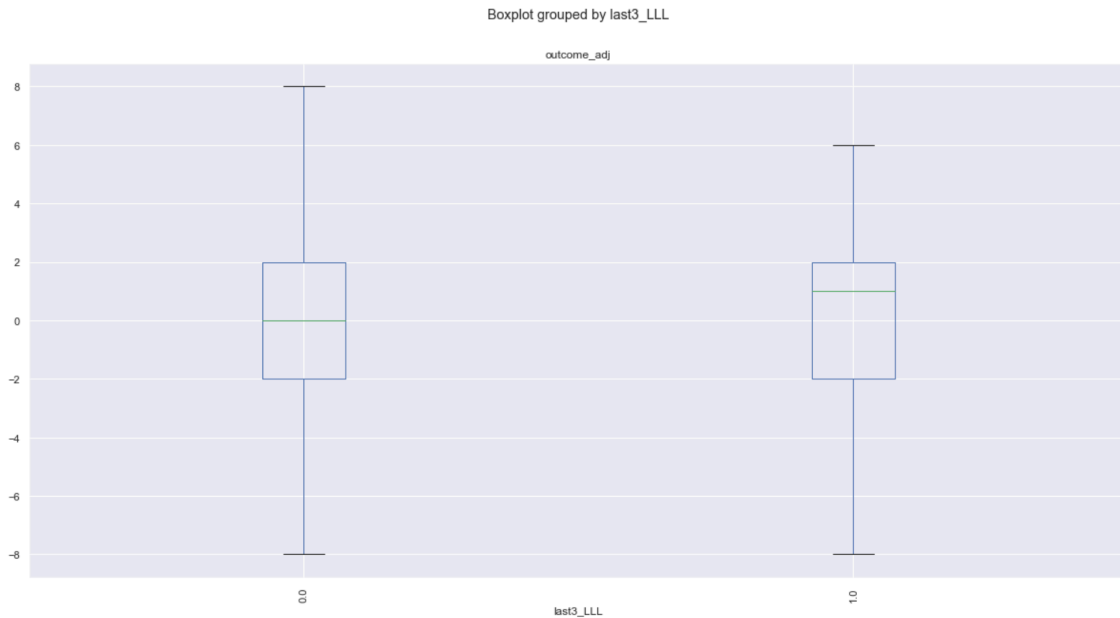
The data points at 0.2 and 0.8 seem to support the hypothesis, with those teams covering and failing to cover, respectively, by nearly 2 touchdowns.

I will continue to isolate and filter the data set to determine if other pre-pick metrics align with these scenarios and try to isolate that data and see if the correlation improves.

Data (one year example):
NFL game data: https://www.pro-football-reference.com/years/2009/games.htm
Pick data: https://www.wunderdog.com/public-consensus/nfl.html?date=09%2F13%2F2009