

What are the odds?

The ultimate guide to naïve sports gambling.

Primary Question: Do naïve sports gambling strategies work?

Sports Gambling is a massive industry with an estimated market capitalization of \$250 billion worldwide, according to Statista. American sports gambling is the fastest growing market segment, increasing from \$20 billion in 2006 to \$40 billion in 2016. Current estimates vary between \$60-73 billion, resulting in an about 10% YoY growth. The upshot? This is a massive industry that is experiencing rapid growth. It's certainly not as secure as the stock market, but it does attract a healthy-sized crowd.

According to thesportsgeek.com, a winning betting percentage of 55% is all that's required to make money betting on sports.

So, can a naïve sports gambler make lots of money? Or would the naïve sports gambler go for broke?

Strategies that we explore to answer this question:

1. What if I always bet on the underdog? (Or the favorite?)
2. What if I always bet on the home team?
3. What if I always bet on the team with the higher win percentage?
4. What's the difference in returns between betting on the money line versus the spread?
5. What if I use different bookies?

The Data: NBA Historical Stats and Betting Data

We used three primary datasets, courtesy of Kaggle.com, to explore our various naïve betting strategies:

1. **nba_betting_money_line.csv** - This file contains information on the money line (i.e. Win/Loss data) odds and returns.
2. **nba_betting_spread.csv** - This file contains information on the spread and spread price. The big difference here is that what matters most is the favorite covering a point spread—it's not enough just to win.
3. **nba_betting_totals.csv** - This file contains game stats on every game that's contained within our money line and spread dataframes.

There were data on ten different bookies. Because bookies all have their own way of determining odds, some bookies may be more appealing for certain naïve betting strategies.

Data Transformation: A time-consuming process

In order to perform our exploratory analysis, we first needed to “massage” our data into a form that made it usable. This first required merging the three data frames into one master data frame, dropping irrelevant columns, and masking our data frame to only include one game per line per bookie.

Working through the logic to create flag columns for easy identification of the winning bets was the most difficult portion of the data transformation workload. After multiple conversations, debates, and pseudo-code jam sessions, we identified which bets would win based on sports odds logic and created flag columns for easy-to-use masks.

Exploratory analysis: Which naïve strategies historically performed the best?

Assumptions: To keep things simple, we only bet \$100 on each bet. If the odds are tied (i.e. no clear favorite or underdog) then we do not bet.

The Underdog

Betting on the underdog every time could result in minimal gains but most likely in significant losses. There were instances of positive returns depending on the bookie that you are looking at, however, those returns were about 0.39% of the total investment—not enough to be regarded as a viable strategy.

The Favorite

Betting on the favorite will almost certainly result in losing money. The returns on each win just aren’t high enough to make up for the losses that you will certainly accrue.

The Home Team Advantage

Assumption: According to our data, the home team wins 59.66% of the time, so there is evidence that this strategy could be fruitful depending on how it is executed. However, the test results were volatile, and returns varied depending on the bookie you chose and whether or not you were betting on the spread. This tactic (thus far) performed the best when betting on the spread and performed poorly when betting on the outright money line.

Betting the Moneyline on the Home Team

Testing this strategy with all bookies combined:

Wins: 59.14% of the time on average

In aggregate, would "win": -\$450,460.52 total

Would require a capital investment of: \$12,157,300 total

Expected value: -\$3.74 per \$100 gambled

Betting the Spread on the Home Team

Testing this strategy with all bookies combined:

Wins: 52.29% of the time on average

In aggregate, would win: \$68,739.03 total

Would require a capital investment of: \$12,157,300 total

Expected value: \$0.39 per \$100 gambled

The Winner

This strategy is the most interesting. We only bet on the team with the higher win percentage. Shockingly, this strategy yielded positive returns over the 12-year aggregate. However, daily returns were highly volatile, suggesting that one could go broke at any point in time during the first couple seasons of gambling on basketball games.

Betting the Moneyline on the Team with the Highest Winning %

Testing this strategy with all bookies combined:

Wins: 71.52% of the time on average

In aggregate, would win: \$1,495,471 total

Would require a capital investment of: \$12,157,300 total

Expected value: \$12.27 per \$100 gambled

Betting the Spread on the Team with the Highest Winning %

Testing this strategy with all bookies combined:

Wins: 54.72% of the time

In aggregate, would win: \$622,746.97

Would require a capital investment of: \$12,157,300

Expected value: \$5.03 per \$100 gambled

Project Outcome: Solution and insights

Sports Gambling is both art and science. Though it is tempting to look at the positive expected values for betting on the team with higher win percentage each time, there is high volatility in that strategy and could result in going broke during any given time interval. One could theoretically win a "large" sum of money with enough capital investment so long as he or she knew when to pull out and stop

gambling. It is important to note that our strategies were tested on every game, which according to many media outlets is a poor idea.

The best naïve sports gambling strategy that we tested was betting on the team with the highest win percentage every time. In aggregate it yielded the most returns and had the highest expected value.

Given the analyses, we would not recommend sports gambling with a naïve approach. Either bet on sports that you know very well and use more of “gut instinct” approach or create models to predict those outcomes. This leads us to our recommended next steps.

Next Steps

For the next steps, we recommend creating two types of models:

- 1. Linear Regression to predict points scored by the home and away team**
 - a. Use this to start testing strategies that use your model to predict spreads
- 2. Classification to predict win or loss**
 - a. Use this to start testing strategies that use your model to predict money lines