Jack Lay
Data Analysis Project

Topic

- I decided to analyze sports betting data for the NBA, in order to see if betting against the spread is actually more profitable than betting on the Moneyline UnderDog. Realistically, the results of this experiment can be relevant to anyone, but the main audience would be those interested in sports betting, or are already involved.

Hypothesis

- The null hypothesis for my experiment is that betting on the spread is in general more profitable than betting on the ML underdog, my hypothesis is that betting on the underdog is more profitable than betting against the spread.

Data Sources

- The first data sets I found were all very large, containing vast amounts of data, but they were not particularly clean. I ended up using the next data set I found. It was clean data from every game, broken up into seasons with point spreads, ML odds, and more statistics that I did not use.
- The biases in these data sets are fairly obvious, as the data I got is derived from Sports Betting sites, there may be an inherent bias towards the house.

Data Analysis

- Variables:
  - For the data i used, the ML and Open Spread as my independent variables, and the net money they generated for each game as my dependent variable.
- Data Visualization:
  - To wrangle this data I used excel as a means of creating columns with meaningful data.
  - If we look at the box plots, we can see that my hypothesis likely was wrong, as across the board on the ML graph, the mean amount earned is $-100. Spread graph shows the average return is to be ~ $100.
- Linear Regression:
  - After performing Linear Regression the following results were found:
    - The correlation between the Open Spread and Net Gain is strong, while the correlation ML and Net Gain could be argued, but is not as strong.
  - Linear Regression between the independent variables revealed that they are not strongly correlated, almost at all.

Summary & Recommendations

- Findings
  - After finding the results of this experiment, I was unable to reject the null hypothesis like I intended to.
- Recommendations
  - Based on the analytical results, my recommendation is different depending on your risk tolerance. If you have lower risk tolerance, you will see more consistent results betting against the spread. But if your risk tolerance is higher, betting on ML underdog will have you see less consistent, but higher yield returns.
  - This is not information that is particularly new, and has been the common consensus in sports betting. The conclusion derived from the data confirms that this is the truth.