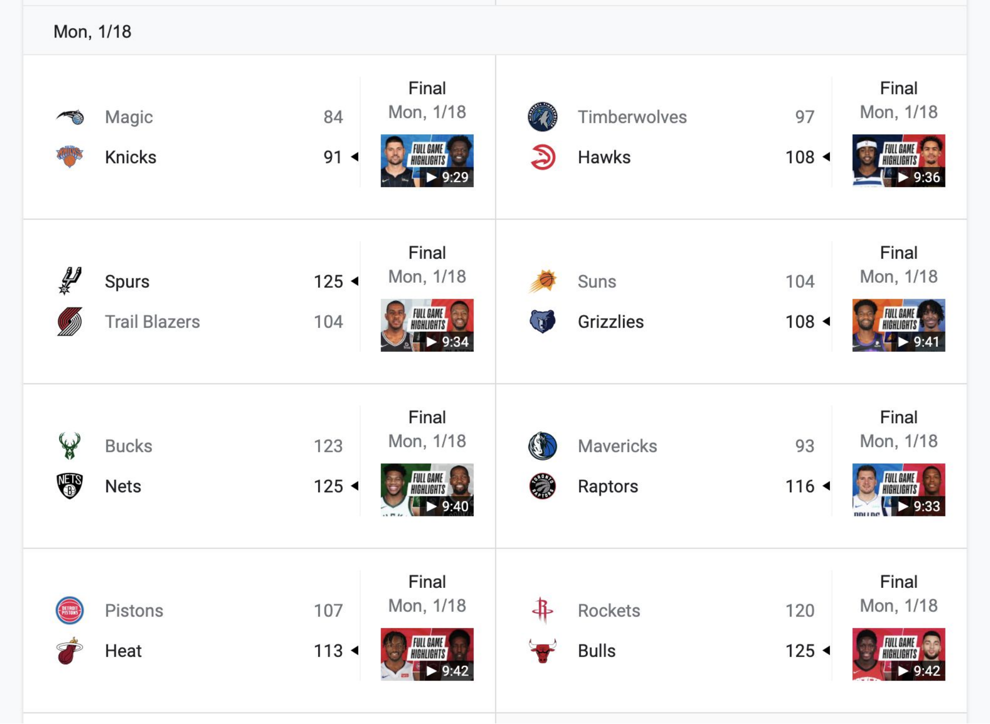
Betting Against the Spread:
NBA Score Prediction

Colin Salama

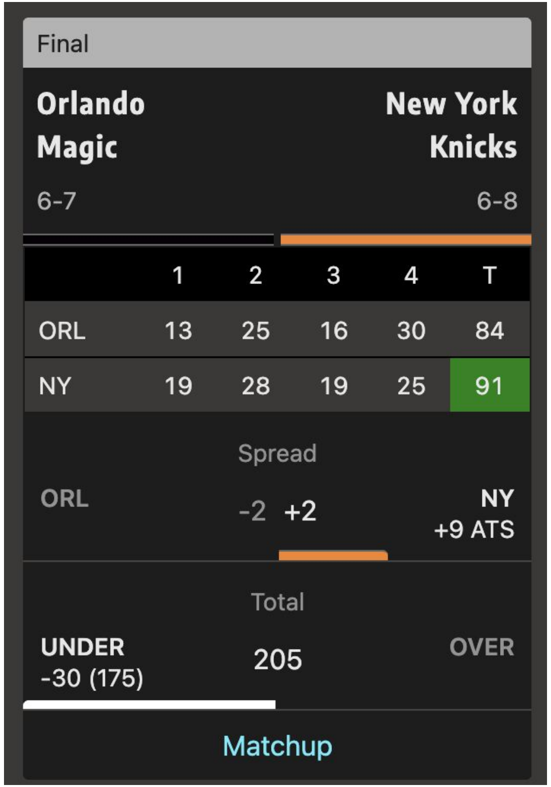
Background/Motivation

- **Sports gambling** is growing due to new government regulations
- **NBA Analytics** are at an all time high - predicting the score is of interest to many
- Could a team's **recent momentum** help in predicting future games?



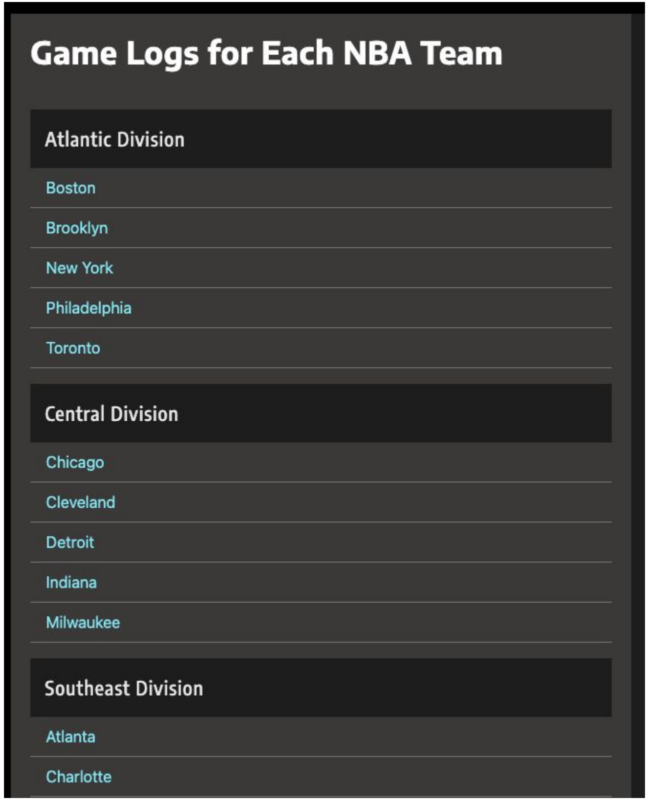
Winning Against the Spread

- Orlando's spread of -2:
  - Because Orlando is favored, they expect that Orlando will score around 2 more points than New York
  - NY wins against the spread here because they scored 9 more points than Orlando… but they only needed to score 1 **fewer** point to win
- Moneyline
  - Depends on the gambling book, but it will generally show the money you win from a bet



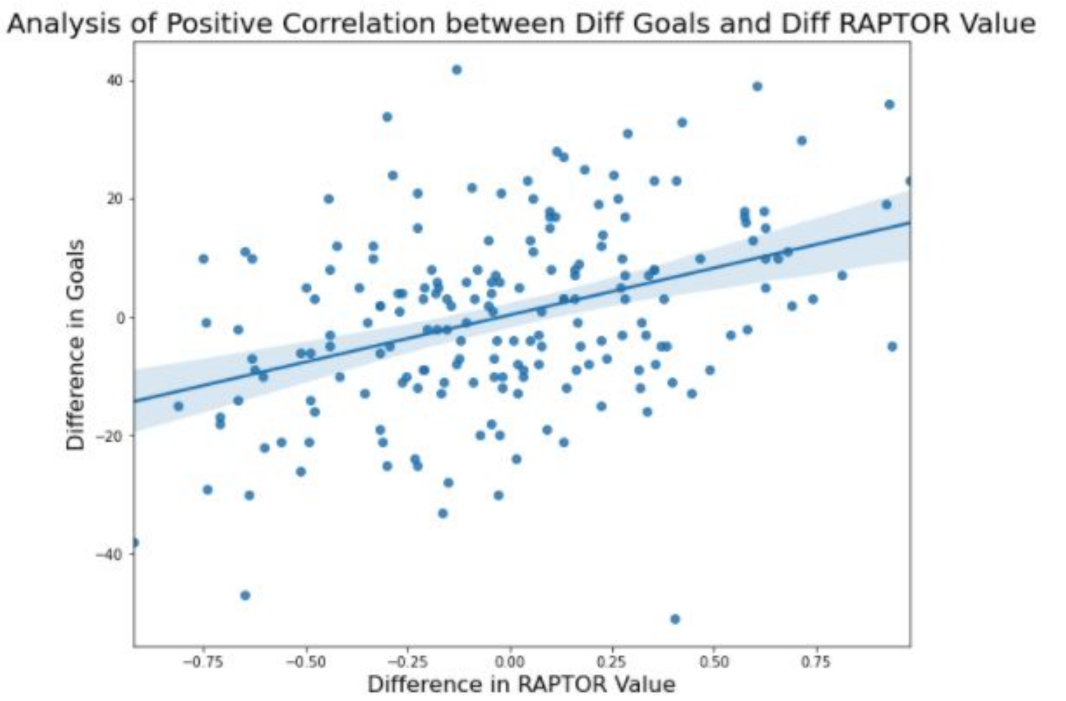
Data Sources

- **Oddsshark.com:** Scraped all Game Data and Spread Data between 2017 and 2021
- **RAPTOR** player value data taken from FiveThirtyEight.com dataset
  - Weighted average of RAPTOR by minutes played calculated to determine overall Team Value

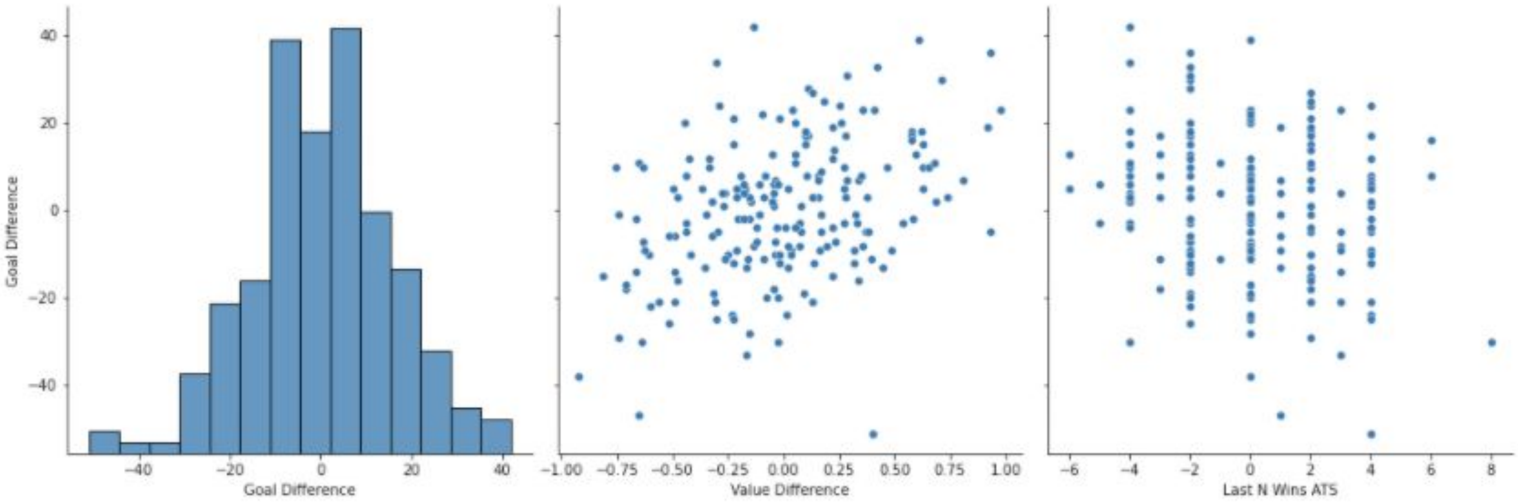


Model

- **Game Data**, 2017-2021
- **RAPTOR**, a positive predictor of Goal Difference
- **Home Team Factor**
- Recent Success **Against the Spread**



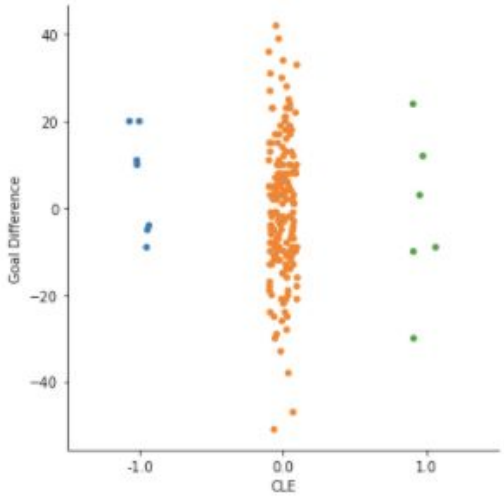
Model - Initial Look



Model - Regression

$$Score\_Diff = \alpha + \beta_1 Value\_Diff + \beta_2 Last\_N\_ATS + \beta_3 Season + \beta_4 ATL + \beta_5 BOS + \ldots + \beta_{33} WAS$$

- **Score Difference** gives the difference between Home team score and Away team score
- **Value Difference** gives the difference between Home Team Value and Away team value
- **Last N ATS** gives the team's record against the spread for their last 5 games.
- Indicator Variable is 1 if the team is at home and -1 if team is away for **every NBA team** to provide a Home Team Factor.
- **Season factor** included all years 2017-2021



Model - Testing and Validation

- R^2 values were fairly low for all models (~20%) due to the high amount of randomness in NBA scores
    - Note, however, that if we know with a higher degree of certainty than the Moneyline implies, we will win in the long run.
- Predictive models considered:
    - **Basic Regression Model**
    - **Polynomial Regression Model**
    - **Ridge Regression**
    - **LASSO Regression**
- 5-Fold Cross-Validation was used to test all models and R^2 values as comparison
- Model was significantly improved when separate model was fit for each season

Conclusion

- **Ridge Regression** model had the strongest R squared score of all methods tested with **~22%**
- Looking only at a **single season (2021)**, the R squared is improved to **~29%**
  - The table to the right gives the coefficient values for each within 2021
- **ATS Variable:** When the Home team has won more games against the spread recently, the model predicts that the Away team will perform better
  - Possibly showing a regression towards the mean

Feature	Coefficient
value_diff	2.57
last_n_ATS_diff	-1.76
ATL	0.29
BOS	-0.25
BRK	0.30
CHA	0.22
CHI	-0.23
CLE	-0.28
DAL	-0.28
DEN	-0.09
DET	-0.09
GSW	0.85
HOU	-0.11
IND	0.39
LAC	1.48
LAL	1.30
MEM	0.20
MIA	-0.31
MIL	1.00
MIN	-1.15
NOP	-0.38
NYK	-0.10
OKC	-1.72
ORL	-1.25
PHI	0.32
PHO	0.31
POR	-0.26
SAC	-1.16
SAS	0.10
TOR	-0.05
UTA	0.52
WAS	-0.00

Future Work

- **Moneyline testing for multiple sports-books**
    - Look at the probability that the score difference is beyond the spread, and if this is greater than the Moneyline-implied probability, bet on this game
- **Logistic Regression Fit**
    - Fitting based on a Win ATS vs Loss ATS may increase the fit