

An Analysis of the NBA for Gambling

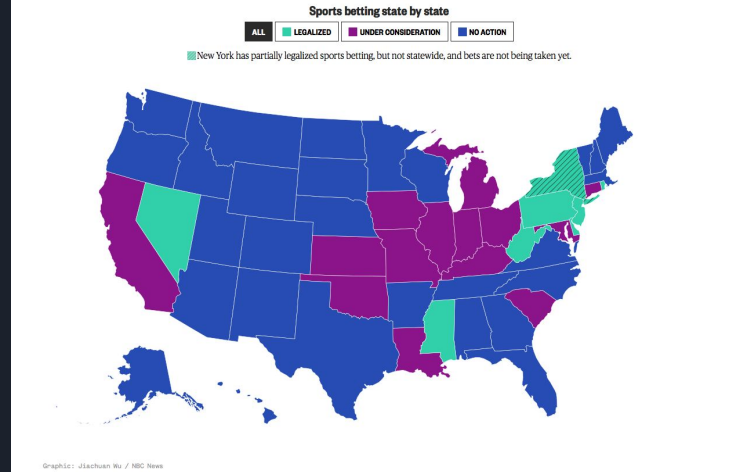


Jeffrey Ebert



Background

- As the amount of states where sports betting becomes legal increases, there are more people gambling throughout the US than ever before
- There are casual bettors who are looking to have fun and throw a few dollars on a game one night and there are serious bettors who are trying to beat the bookies and win a significant amount of money on sports
- With more people betting on sports, there is an increase in demand for information on sports that could give gamblers a potential edge in understanding game trends in an attempt to make money
- Basketball and the NBA have continued to be one of the most popular sports to bet on as the season lasts from October to June.





Objective

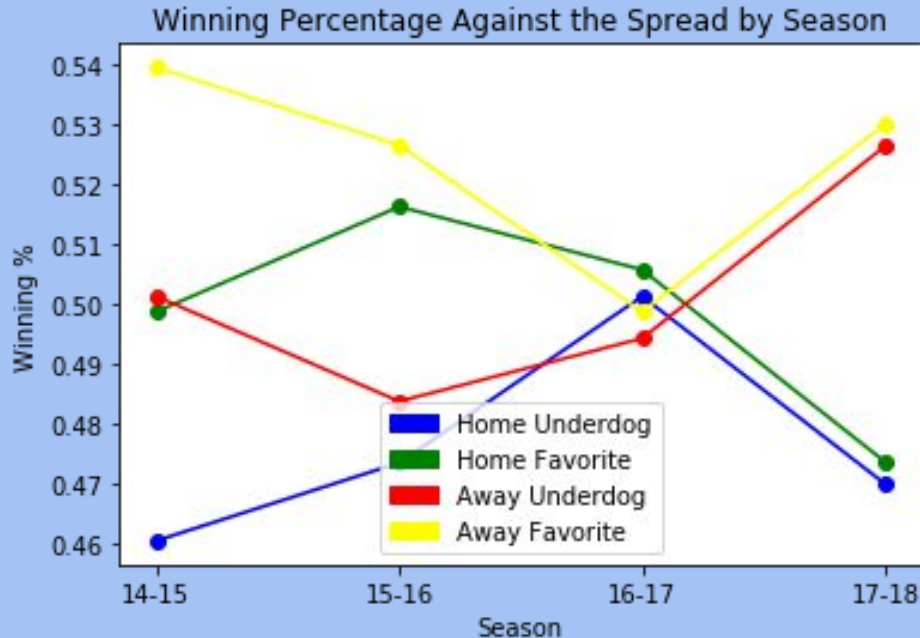
- Evaluating important trends and factors throughout an NBA season that can provide insight on whether a certain team will win or lose a game and cover the spread
- Determining what are the most significant game and season long statistics for a team in winning a game and using machine learning models to predict the outcome of a game based on those statistics



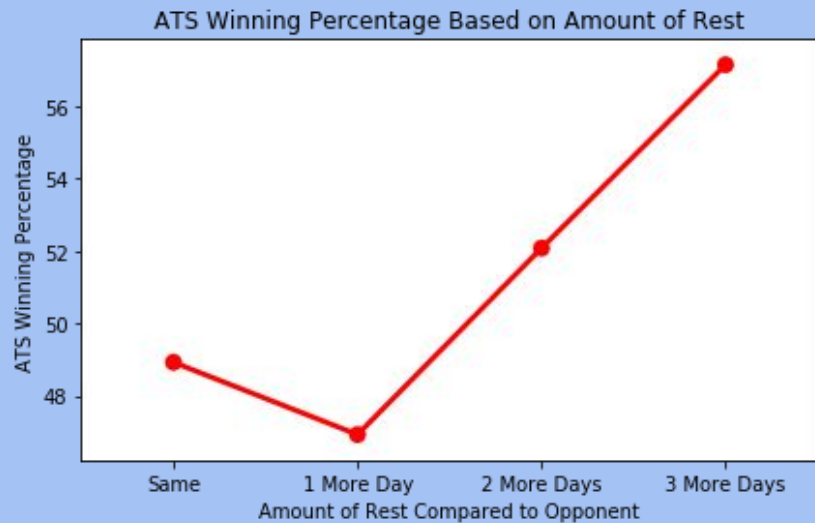
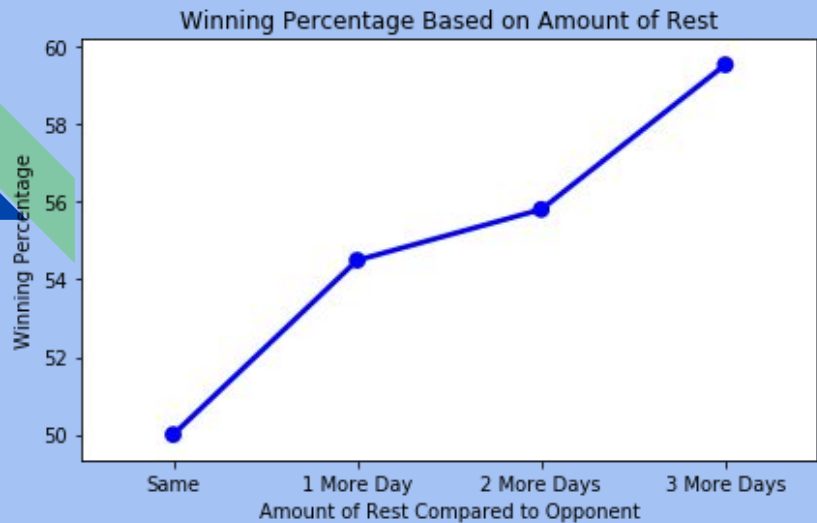
Data

- All teams, schedules and game scores were uploaded via a web scraper from ESPN.com
- The corresponding statistics and official lines for each game were uploaded from public datasets
- The data used consists of every regular season NBA game played during the 2014-2015, 2015-2016, 2016-2017, and 2017-2018 seasons
- This information was used to form various features about a game such as if a team is playing at home or on the road, records, points differentials, winning streaks, amount of rest and game variables

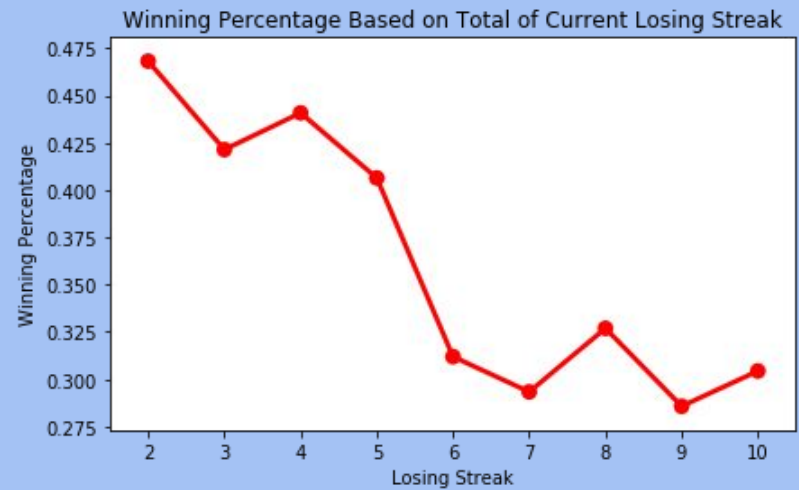
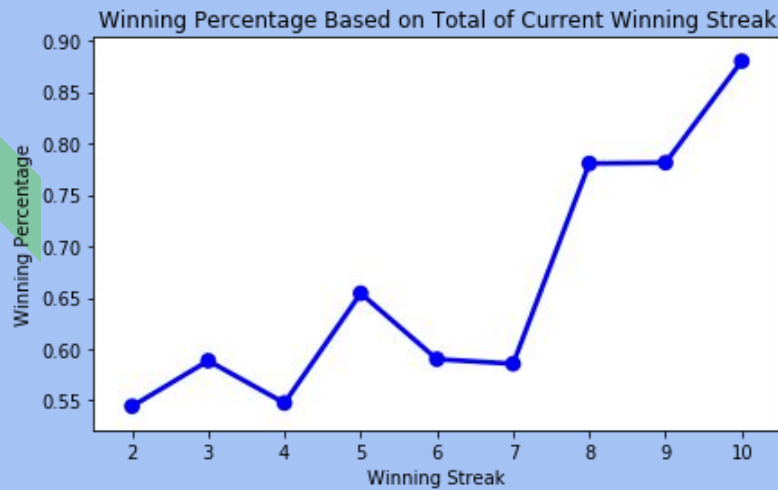
Trends



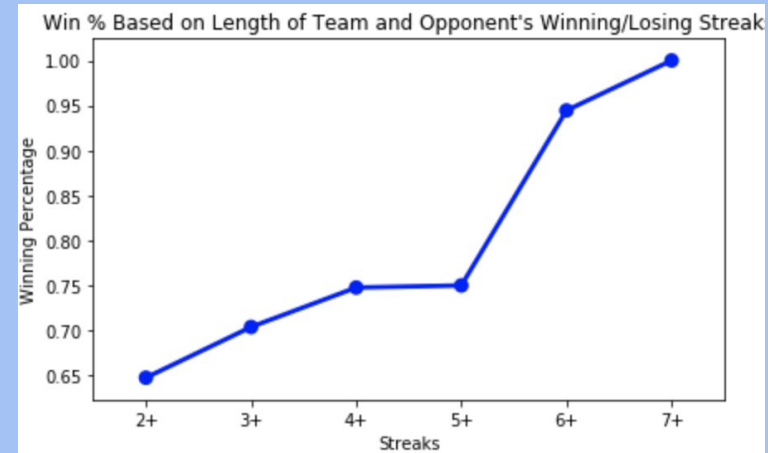
- From season to season, depending on if a team is at home, on the road, and an underdog or a favorite does not provide much information in predicting if a team will win or not and cover the spread.
- This does not mean certain teams do not provide definitive trends depending on these characteristics from year to year and within a season.
- Because these broad characteristics of a team are not enough to gain insight, more attributes of a game need to be analyzed.



- A trend that reveals a high correlation with winning and winning against the spread is the amount of rest a team is on compared to their opponent. As the difference in days off between two teams increases so do these winning percentages. These percentages can become even more dramatic during long stretches of games for certain teams.



- Another logical and definitive trend of winning teams is their recent success compared to their opponent's. As a team's winning streak increases they are more likely to win their next game and vice versa for a team on an increasing losing streak
- These percentages become more drastic for teams with opposing streaks. If a team on winning streak is facing an opponent on a losing streak, they are more likely to win as these streaks diverge and lengthen



Statistics

- This matrix represents the correlation each seasonal average statistic has with winning and the other stats

	win	A_FGM	A_FGA	A_FG%	A_3PM	A_3PA	A_3P%	A_FTM	A_FTA	A_FT%	A_OREB	A_REB	A_AST	A_STL	A_BLK	A_TO	A_PF
win	1	0.11	-0.013	0.15	0.1	0.079	0.095	0.033	0.016	0.051	-0.024	0.042	0.12	0.065	0.066	-0.048	-0.055
A_FGM	0.11	1	0.59	0.77	0.38	0.27	0.42	-0.085	-0.16	0.19	-0.032	0.25	0.67	0.21	0.17	-0.0069	-0.092
A_FGA	-0.013	0.59	1	-0.068	0.24	0.29	-0.043	-0.14	-0.14	-0.017	0.37	0.54	0.23	0.11	0.019	-0.028	0.027
A_FG%	0.15	0.77	-0.068	1	0.28	0.098	0.56	0.0025	-0.092	0.25	-0.33	-0.11	0.63	0.16	0.19	0.0072	-0.14
A_3PM	0.1	0.38	0.24	0.28	1	0.94	0.52	-0.086	-0.12	0.092	-0.26	0.063	0.38	0.12	0.027	0.088	-0.073
A_3PA	0.079	0.27	0.29	0.098	0.94	1	0.22	-0.064	-0.069	0.024	-0.21	0.086	0.27	0.12	0.0043	0.11	-0.03
A_3P%	0.095	0.42	-0.043	0.56	0.52	0.22	1	-0.11	-0.19	0.2	-0.24	-0.045	0.41	0.044	0.064	-0.017	-0.13
A_FTM	0.033	-0.085	-0.14	0.0025	-0.086	-0.064	-0.11	1	0.92	0.28	0.17	0.14	-0.14	0.063	0.025	0.0094	0.23
A_FTA	0.016	-0.16	-0.14	-0.092	-0.12	-0.069	-0.19	0.92	1	-0.098	0.26	0.19	-0.2	0.064	0.021	0.068	0.25
A_FT%	0.051	0.19	-0.017	0.25	0.092	0.024	0.2	0.28	-0.098	1	-0.23	-0.13	0.16	0.021	0.022	-0.15	-0.048
A_OREB	-0.024	-0.032	0.37	-0.33	-0.26	-0.21	-0.24	0.17	0.26	-0.23	1	0.57	-0.22	0.0024	-0.0049	0.13	0.15
A_REB	0.042	0.25	0.54	-0.11	0.063	0.086	-0.045	0.14	0.19	-0.13	0.57	1	0.043	-0.19	0.23	0.18	-0.046
A_AST	0.12	0.67	0.23	0.63	0.38	0.27	0.41	-0.14	-0.2	0.16	-0.22	0.043	1	0.26	0.25	0.11	-0.14
A_STL	0.065	0.21	0.11	0.16	0.12	0.12	0.044	0.063	0.064	0.021	0.0024	-0.19	0.26	1	0.033	0.26	0.16
A_BLK	0.066	0.17	0.019	0.19	0.027	0.0043	0.064	0.025	0.021	0.022	-0.0049	0.23	0.25	0.033	1	0.11	-0.0068
A_TO	-0.048	-0.0069	-0.028	0.0072	0.088	0.11	-0.017	0.0094	0.068	-0.15	0.13	0.18	0.11	0.26	0.11	1	0.35
A_PF	-0.055	-0.092	0.027	-0.14	-0.073	-0.03	-0.13	0.23	0.25	-0.048	0.15	-0.046	-0.14	0.16	-0.0068	0.35	1

- Based on the matrix the statistics that are most correlated with winning are averaging a large amount of assists and a high field goal percentage
- Using these basic statistics and other game features such as whether a team is at home, who is playing, and who they are playing against, a model was created using a Gradient Boosting Classifier that predicted the winner of any game 66% of the time
- While this number is not outstanding, it can be improved upon by using more advanced statistics and data