Predicting NBA Outcomes

Rosario Chiovaro, Nathan Davis, Daniel Eschman, Jill Mantel, Vishnu Nistala

Motivation / Introduction

Success in the playoffs in a sports league such as the NBA is often random and highly unpredictable – and requires making it there in the first place. Many decisions made by NBA teams would be made simpler if the teams' decision makers could predict to a high degree of accuracy how many regular season games their team would win.

Approaches

In contrast to existing analysis which is primarily limited to a focus on likelihood of making the playoffs or winning the title, our approach focuses on projecting team rank for a given season, as opposed to success in the playoffs or championship games. Additionally, our approach utilizes a straight-forward linear regression model which was trained on numerous factors in team statistics determined to be most indicative of success. This model has the advantage of being less complex than more common ensemble methods while also offering a more explainable relationship between attributes and responses.

Data

Data was pulled from the Basketball Reference branch of the Sports Reference API, which grabs data from basketball-reference.com. Upon completing data cleaning and manipulation, we obtained finalized data frames for data analysis, which featured 90 rows (three seasons, with rows for each of the 30 teams) and 54 columns of statistics spanning a huge variety of metrics.

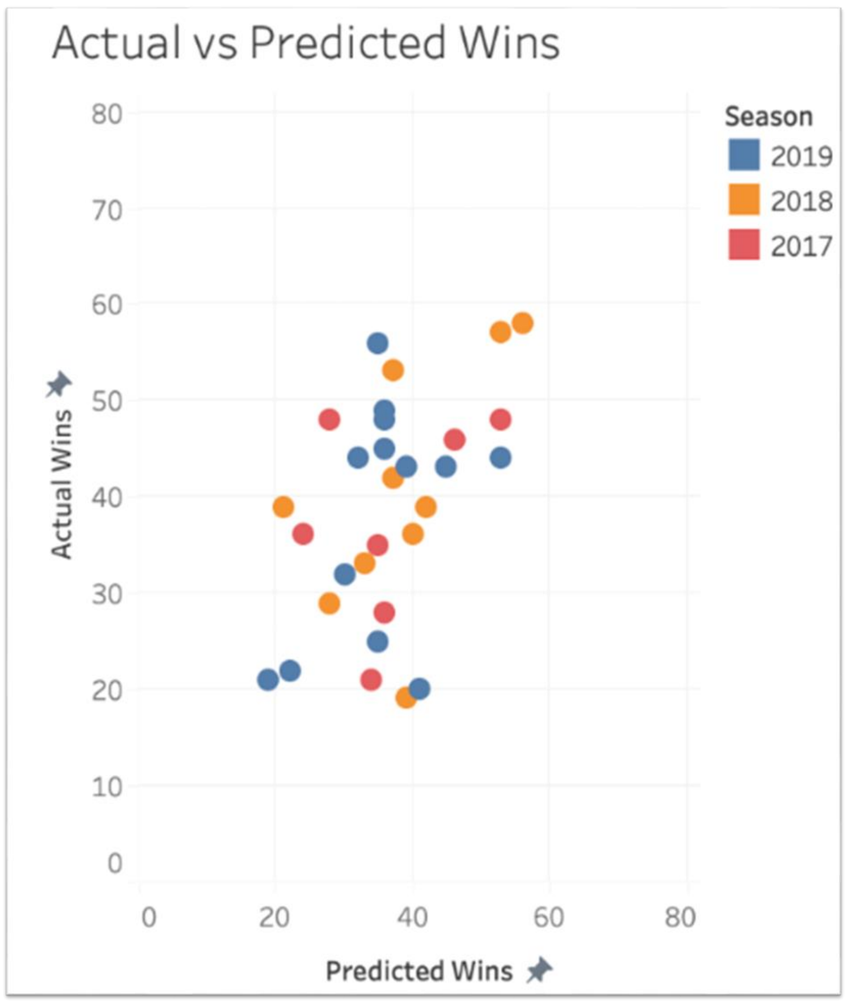
	abbreviation	assists	blocks	defensive_rebounds
GSW	GSW	2491	555	2873
HOU	HOU	2070	352	2746
DEN	DEN	2077	323	2837
CLE	CLE	1858	327	2821
WAS	WAS	1956	336	2671
...	...	...	...	...
MIA	MIA	1991	448	2879
CHI	CHI	1796	351	2799
NYK	NYK	1646	422	2811
CLE	CLE	1698	195	2619
MEM	MEM	1963	448	2703

120 rows × 48 columns

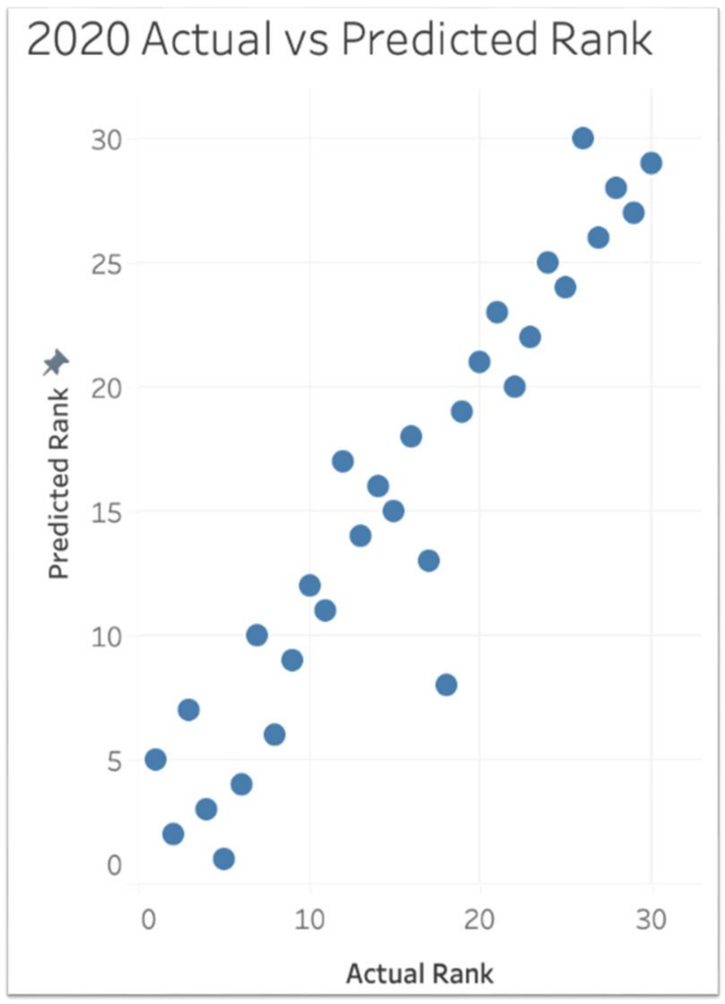
Experiments/Results

Our model was initially fit on a subset of the data using linear regression (2017 - 2019 seasons) and the accuracy was evaluated using the 2020 season to compare with the real-world results. However, it was found that roster turnover between seasons makes win prediction in this way impractical and highly inaccurate.

However, predicting the final finishing rank of each team proved to be more realistic: approx. 72% of team rank predictions were within three ranks of the actual finishing rank.



Our model proved to be inaccurate when predicting wins – while some teams were predicted reasonably accurately, many others were far off.



Using our model to predict a team's final finishing rank proved to be much more accurate, with only a handful of serious outliers.