NBA Draft Analysis

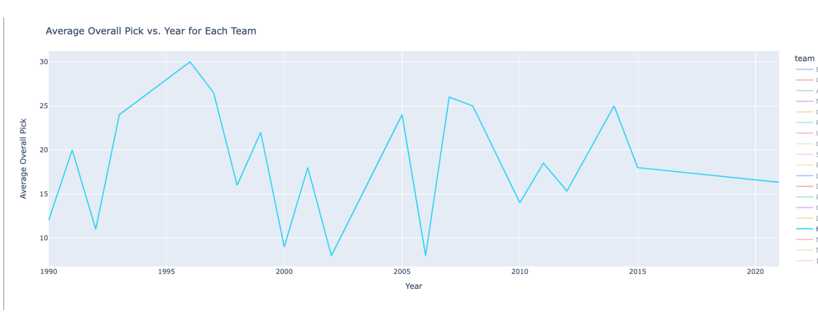
By: Andrew Ng

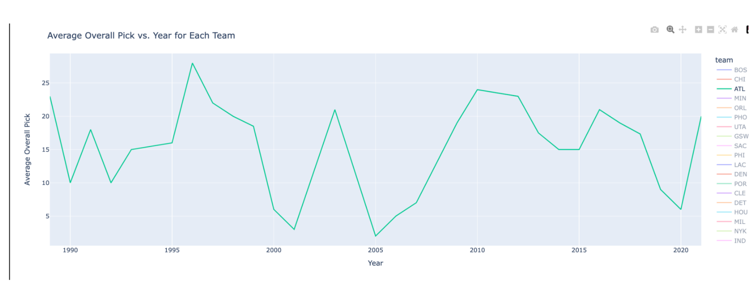
Introduction:

The NBA Draft is an eagerly awaited yearly event that offers teams the opportunity to recruit exceptional talent with the aim of building a strong roster in the hopes of attaining the highly coveted championship. For franchises that have struggled, the draft provides a great chance to even the odds by acquiring promising young players who can form a strong core for future success. In addition, it also presents an opportunity to target specific players who can fill crucial roles within their team. All together, the draft provides exciting opportunities and risky possibilities for teams across the league. This report aims to determine the value of draft position by production of players drafted with each pick in the NBA draft and finding over performing and underperforming NBA teams and college teams, draft-wise and prospect-wise respectively.

Data Comprehension:

For this paper, I investigated a Kaggle NBA draft dataset, which contains draft data from 1989 to 2021. The dataset contained 24 distinct features, ranging from pick information to player stats. To learn more about this dataset, I began simple exploratory data analysis. First, I looked at which NBA teams drafted the most players who went to Duke and were drafted in or before the 2000 draft. After manipulating the data, I found that the Timberwolves, Mavericks, and Suns drafted the most, each with two. Next, I looked at which NBA team(s) has drafted the most players who have a first name that begins with D and were drafted in an even year draft (1990, 1992, 1994, ...). Again, after manipulating the data, I found that the Bucks, Celtics, and Sonics (inactive) all drafted the most, each with seven. Finally, I looked to describe the relationship between a team's first round pick slot in one year with their first-round pick slot in the subsequent year. To tackle this, I first attempted to plot first round picks of each team versus draft year. However, this became problematic for cases where teams had multiple first round picks in a common draft year. In order to combat this, I calculated the average pick number for teams that had multiple first round picks in the same draft year. Following, I plotted the average pick vs year for each team. Here are some figures below.





Just by eyeballing these images, we can observe stand-out relationships. First, we see that when teams obtain an average pick of 10 or below, teams commonly obtain a much higher average pick the following year. In addition, we see that a sharp increase or worsening in average pick from previous year leads to sharp decrease or improvement of average pick the subsequent year.

Analytical Acumen

With the given task of analyzing draft position value and team success/deficiencies compared to expectation, I first tackled finding a methodology of valuing each pick slot. In order to do so, I had to figure out which feature in the dataset I wanted to consider as value. In my methodology, I assign value to each pick based on historical player win shares. Initially, I wanted to value each pick based on win shares and player efficient rating (PER). However, the given dataset missed many critical statistics, such as blocks and turnovers, that would be needed to assign each player a fair PER rating. Overall, the end goal in basketball is to win games (to win championships), so choosing win shares as my value of each pick was intuitive. I created a regression model between pick number and win shares, which resulted in a polynomial relationship. Below is the value of each pick from the model:

overall_pick	draft_pick_value
1	1
2	0.901999
3	0.81561
4	0.739704
5	0.673222
6	0.615178
7	0.564654
8	0.520799
9	0.482824
10	0.450005
11	0.421675
12	0.397224
13	0.376097
14	0.357793
15	0.341859
16	0.327891
17	0.315529
18	0.304458
19	0.294401
20	0.285122
21	0.27642
22	0.268129
23	0.260111
24	0.252261
25	0.244498
26	0.236767
27	0.229035
28	0.221288
29	0.213529
30	0.205778
31	0.198066
32	0.190435
33	0.182936
34	0.175624
35	0.168558
36	0.161801
37	0.15541
38	0.149441
39	0.143945
40	0.138962
41	0.134526
42	0.130652
43	0.127346
44	0.124592
45	0.122356
46	0.120583
47	0.11919
48	0.11807
49	0.117087
50	0.116071
51	0.114822
52	0.113099
53	0.110628
54	0.107089
55	0.102122
56	0.0953204
57	0.0862307
58	0.0743484
59	0.0591168
60	0.0399243

From this, we can see just how valuable obtaining a high pick in the draft. According to our model, even obtaining the second pick is nearly 10% worse in value than obtaining the coveted first pick.

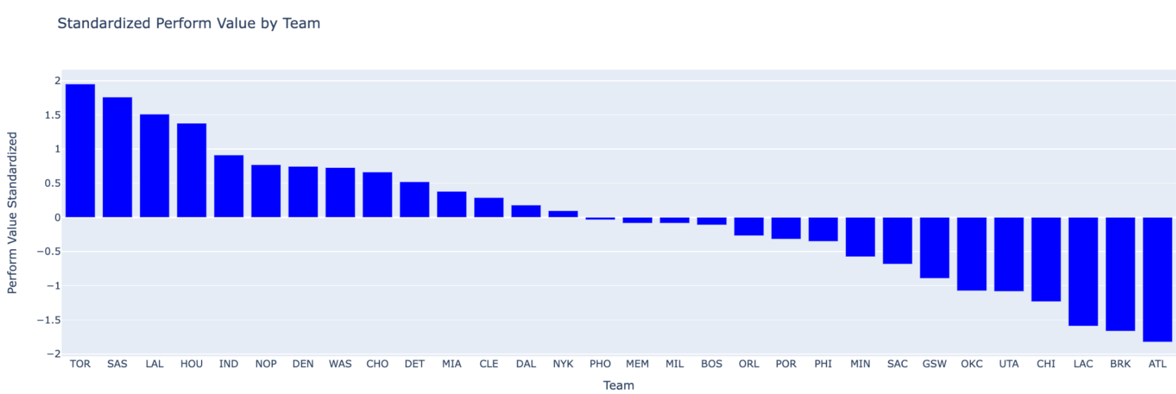
The next step was now to find which NBA teams have over or underperformed the most drafting-wise. In addition, which college teams have had players that have outperformed expectations the most.

In order to tackle both of these questions, I needed to create a method in assigning each player a performance value. The method I decided to implement was to compare each player's feature against the mean or expected value of features where all players were drafted at the same pick slot as the selected player. First, I found the expected value for each feature in the dataset, conditioning on pick number. For example, in our dataset, first overall picks have an expected number of games of 630. Next, I wrote a helper function that would aid my method. In the implementation of my method, I assigned positive scores if the player's stat was above the expected value of the same stat and negative scores vice versa. In addition, I assigned heavier positive or negative scores if the player's stat was at least one standard deviation above or below the expected value respectively. Finally, I applied this function across my initial dataset to output a new feature named performance_value. Let's take a look a couple of players with their performance value:

overall_pick	player	Perform Value
1	Allen Iverson	10
1	Kwame Brown	-20
1	LeBron James	14
1	Anthony Bennett	-26

Here are four former first overall picks, two of which former NBA MVPs and two others considered by many as busts. As we can see, each performance value score seems appropriate for each player.

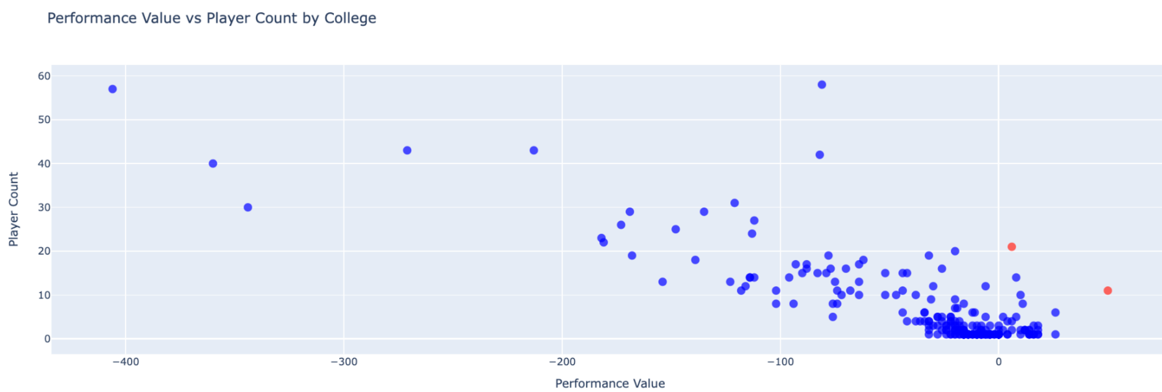
Once I obtained a performance value for each player, I then aggregated the performance value score for all the players drafted by each team. Finally, I standardized each team's score.



Notice inactive teams in the dataset are not included. Rather, I added their respective score to their active franchises. For example, the Bullets's' score was added to the Wizards's score. From this, I found that the Toronto Raptors were the best drafting team in the NBA, followed by the San Antonio Spurs and Los Angeles Lakers. This comes as no surprise, as all of these four franchises have won at least one championship within the time period encompassed by the dataset. On the other side of the coin, I found that the Atlanta Hawks were the worst drafting team, followed by the Brooklyn Nets and Los Angeles Clippers.

Unsurprisingly, none of these teams have won a championship within the time frame of the dataset.

To find which college teams have had players that have outperformed expectations the most, I aggregate the performance value of each player grouped on common college teams. Then, I plotted the total performance value for each college team versus the number of players drafted from each college.



In accordance with our metrics, an outperforming college team would have sent many players with a high aggregate performance value. This means we would like to observe teams who are closest to the top-right of our plot. Looking at our plot, we observe two standout candidates, each with at least ten drafted players and positive aggregate performance values: Florida and Wake Forest. Both of these programs have been able to obtain positive aggregate performance values, while also sending well above average number of players to the league.

Next Steps

While this report is quite comprehensive, there is abundant room for improvement. First, not being able to access important defensive statistics is troublesome, especially in the process of accurate valuation of players and pick slots. In continuation, while my performance value of each team was relatively accurate, there were some inaccuracies. For example, my methodology ended up with the Golden State Warriors and Chicago Bulls being in the bottom 30th percentile in terms of drafting. However, within the time span of the dataset, each team has won four and six championships respectively. Moving forward, I would research more methodologies about pick valuation and team drafting skill based on more descriptive information. For example, while the first overall pick might be a bust, if the player was the consensus pick, then this should not be considered as a bad pick.