Part 1: Introduction and Research Questions

In the NBA, player salaries are an important consideration when forecasting contracts extensions and free agent signings. Our project aims to evaluate the best "value players" in the NBA over the past ten seasons. Given this, we want to determine the most productive players relative to their salaries each year and look at trends in productivity compared to salary. We define "value players" as those with the highest productivity per percentage of total salary cap for their team. We will answer this research question by developing metrics for productivity based on player statistics, such as plus-minus rating, PER (player efficiency rating) and NET rating. The ratings are calculated as weighted sums of various offensive and defensive statistics, and are considered meaningful statistics by the basketball analytics community. The project required substantial analysis because we needed to calculate various metrics from the raw data and compare it between many players over multiple seasons. Sports analytics is a very relevant field of data science, with huge amounts of money and resources being allocated to it every year. With the results of this study, we hope to find which players were the most "valuable," and we hope to find the best metric for this value.

Part 2: Summary of Results

In order to evaluate the productivity of each player, we had to develop different metrics in order to compare player value. Each of our 3 metrics valued players differently based on their stats and their salary. We determined that eABG, a metric we developed based on a player's NET rating and the proportion of the total season's minutes the player played, was the most effective measurement. In order to get a more accurate picture of productivity, we adjusted our metrics in order to look at the player's salary as a proportion of their team's total salary that year. Once the most productive players in the last year in our data set were determined, Duncan Robinson was found to be the most productive player in 2018-2019, and in the last 20 years in fact.

Part 3: Data Sources

We acquired our player data from the website Kaggle.com, which has numerous NBA related data sets available. The ones we used are historical player statistics and player salaries, as these will be the two factors in determining how productive the players were compared to their contract size. The dataset with player stats from 2000-2020 were found at https://www.kaggle.com/justinas/nba-players-data. These stats were compiled using NBA Stats API, which pulls stats from NBA.com stats. This dataset contains all basic stats and the most common advanced stats for each player. The salary data for each player was found at https://www.kaggle.com/hrfang1995/nba-salaries-by-players-of-season-2000-to-2019. These data were scraped from https://hoopshype.com/salaries/players/ and contains salary data for most players from 2000-2019. In order to verify the correctness of the salary data, we compared the salary data with a data set found at http://www.celticshub.com/2017/12/07/nba-player-salaries-

1991-2017/, which is salary data sourced from the Dallas Morning News. This allowed us to verify that our player salary information was accurate.

For our team salaries, we also used salary data from http://www.celticshub.com/2017/12/07/nba-player-salaries-1991-2017/, from before and corroborated this data with data that we scraped from https://hoopshype.com/salaries/.

Part 4: Results and Methods

In order to determine the productivity of each player, we made an efficiency statistic called Arbitrary Basketball Grade (ABG). The ABG for each player was calculated using the equation $ABG = NET \times \frac{GP}{82} \times USG\%$, where NET is the player's NET rating calculated in the manner described in the introduction, GP is the number of games played, USG% is the average percentage of each game the player played over the course of the year. This was done to reward players who played more often as opposed to injured players or players who were productive on very limited minutes. To take into account player salaries, each player's ABG was divided by their salary to give the effective ABG (eABG). We decided to ignore players that did not have salary information in either salary data set used, since these players were mostly irrelevant in terms of production upon further inspection.

Once each player's eABG was calculated, we found the 5 most productive players each year. Table 1 shows the very most productive players for each year between 2000-2019. From this analysis, one trend we noticed was that most of the players who were the most efficient were either drafted in the second round of their respective draft year or went undrafted. This made sense to us because these players tend to have much smaller contracts than players drafted in the first round, but in order to quantitatively determine if this was the case, we conducted a hypothesis test at the significance level of 0.05 to see if players drafted in the first round had lower eABGs compared to players drafted in the second round or who were undrafted. We also wanted to control for the number of years played for each player, so we conducted the hypothesis test on first-year players drafted in the first round and first-year players not drafted in the first round, on second-year players drafted in the first round and not drafted in the first round, etc. From these results there is evidence to suggest that non-first round players were more productive in their first years, but interestingly the data suggest that first-rounders were more productive from years 3-5. This is interesting given that the most productive players were generally first-rounders, but this is likely because the average first rounder likely has more potential than the average non-first rounder, allowing them to grow into more productive players on average after their first few years in the league. We determined the highest-ranked player for each year from 2000-2019, as shown in Figure 1. Out of curiosity, we also looked at eABG trends for players who had exceptionally high single-season eABG statistics and for the best players in the league. The trends for these players shown in Figures 2-4 are very interesting and should be investigated further.

	player_name	Year
521	Dirk Nowitzki	2000
64	Eduardo Najera	2001
108	Dan Langhi	2002
26	Devin Brown	2003
87	Udonis Haslem	2004
48	Awvee Storey	2005
211	Leandro Barbosa	2006
125	Ronnie Price	2007
148	Thomas Gardner	2008
292	Matt Barnes	2009
22	Mike Harris	2010
3	Danny Green	2011
37	Malcolm Thomas	2012
117	Patrick Beverley	2013
77	Draymond Green	2014
76	Draymond Green	2015
144	Jonathon Simmons	2016
104	Fred VanVleet	2017
19	Monte Morris	2018
0	Alex Caruso	2019

Figure 1. The highest rated players by eABG in each year since 2000.

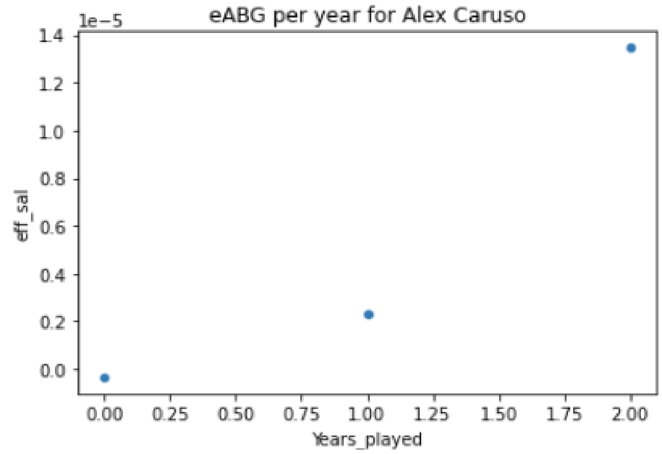

Figure 2. eABG scores for Alex Caruso in each of his NBA seasons.

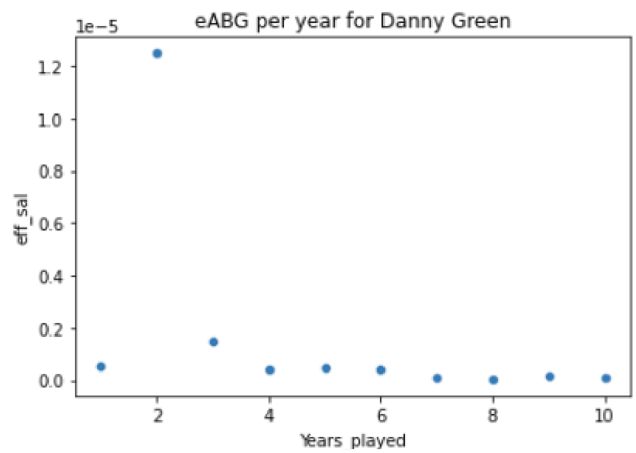

Figure 3. eABG scores for Danny Green in each of his NBA seasons.

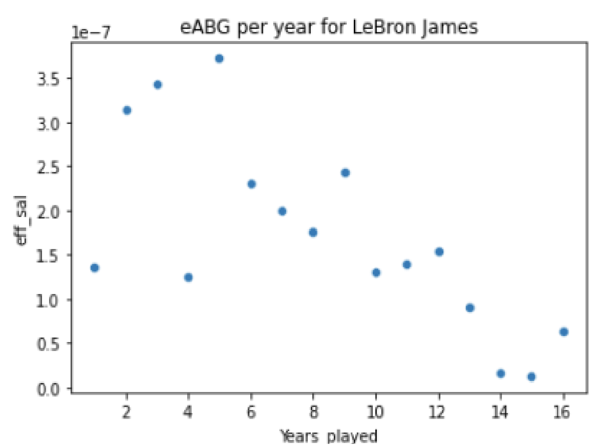

Figure 4. eABG scores for LeBron James in each of his NBA seasons.

Our eABG was just one metric to test productivity. Another common measurement of player performance is win share (WS). WS essentially measures how many team wins the player contributed to during the season, with certain WS metrics being calculated slightly differently than others. The WS metric used was the Basketball-Reference metric, which awards 1 team win share for each team win during the season. Additionally, the WS measurement itself has different weightings than the original WS metric developed by Bill James. Essentially, if a team wins 50 games in the season, then the team will have 50 available win shares, and the WS for each player is the number of wins in the season they are responsible for out of these 50 wins. In our effective win share metric, we used the same treatment as in eABG to account for the amount of time each player played and then divided that value by the salary earned by the player during the season (eWS). In our preliminary testing of the WS metric, we looked at the top 5 players in eWS in the past 20 years (Figure 5).

	Year	Player	team_abbreviation	G	Salaries	eff_WSp48
3158	2007.0	Von Wafer	POR	1.0	24300.0	1.316589e-07
1422	2003.0	Paul Shirley	CHI	2.0	29832.0	9.912575e-08
5813	2013.0	Chris Wright	MIL	3.0	27859.0	3.084046e-08
7001	2016.0	Boban Marjanovic	DET	54.0	1200000.0	1.437530e-08
3437	2008.0	Carl Landry	HOU	42.0	427163.0	1.359758e-08

Figure 5. The top 5 players in eWS in the past 20 years, along with data for games played and salaries.

The top 3 players in eWS since 2000 all played in less than 5 games with extremely low salaries. This presented a problem with the current calculation for eWS. Due to the nature of how WS is calculated, the range of WS values only ranges between approximately 0-20, whereas the salaries can vary by several degrees of magnitudes. In order to alleviate this, we decided to tried to calculate eWS using the following formula: $eWS = \frac{WS \times \frac{GP}{82} \times USG\%}{log(salary)}$. This resulted in a very different top 5 list containing 3 of the best players in the league (Figure 6). This was much different than the previous eWS or eABG calculations due to the WS and minutes played being much more impactful using this equation after using the log(salary). Although these results were much different than those from the eABG calculations, we were still quite happy with these results, since even though these players had very high salaries, it could be argued that their production is well worth it.

	Year	Player	team_abbreviation	G	Salaries	eff_WSp48
3842	2009.0	LeBron James	CLE	81.0	14410581.0	0.002018
5589	2013.0	LeBron James	MIA	76.0	17545000.0	0.001780
6830	2016.0	Stephen Curry	GSW	79.0	11370786.0	0.001751
5509	2013.0	Kevin Durant	OKC	81.0	17548838.0	0.001640
6383	2015.0	Stephen Curry	GSW	80.0	10629213.0	0.001601

Figure 6. The top 5 players in eWS in the past 20 years once log(salary) is used in the calculation.

We then calculated the highest eWS players in each year between 2000-2017 (Figure 7). All the players in the table are all-time greats, so these values make sense. In fact, half of the players in the table won the MVP award the same year that they topped eWS in the NBA. Thus, although this metric emphasizes performance and consistency in minutes played more than eABG, it is still useful in highlighting the productivity of NBA superstars.

	player_name	Year
62	Vince Carter	2000.0
632	Dirk Nowitzki	2001.0
861	Tim Duncan	2002.0
1345	Tracy McGrady	2003.0
1643	Kevin Garnett	2004.0
2250	Dirk Nowitzki	2005.0
2661	Dirk Nowitzki	2006.0
3071	Dirk Nowitzki	2007.0
3503	Chris Paul	2008.0
3842	LeBron James	2009.0
4280	LeBron James	2010.0
4763	LeBron James	2011.0
5175	LeBron James	2012.0
5589	LeBron James	2013.0
5967	Kevin Durant	2014.0
6383	Stephen Curry	2015.0
6830	Stephen Curry	2016.0
7333	James Harden	2017.0

Figure 7. The highest rated players by eWS each year since 2000.

One final metric that was tested was player efficiency rating (PER). This metric attempts to evaluate players holistically and distill all of their contributions into a singular value. The distribution of PER scores is skewed right, with the mean at 15 and with small increases in PER indicating relatively large differences in performance. Because of this, we wanted to preserve the fact that higher PERs are much, much more impactful; a PER of 30 does not mean that they are only twice as good as the average player, it means that they are the clear MVP frontrunner. In order to do this, the true PER of a player was determined to be $100^{3\times0.01\times PER}$ $ePER = \frac{100^{3\times0.01\times PER}}{salary}$. This expression was determined by trial and error until the true PER values of 15 and 30 reflected our interpretations of productivity differences between the average rotation player and an MVP candidate. We decided not to include usage statistics in the numerator since ePER varies exponentially with PER, and players with already high PER will have very high usage rates as well. The players with the highest ePERs in each year are shown below in Figure 8.

In order to compare how each of the three metrics valued player salaries, we plotted the salaries of the highest rated players for that metric by year (Figures 9-11). ePER favored the lower played players the most, whereas eWS favored the highest paid players. eABG served as a good middle ground between the two, with some of the top players having very low salaries, and some of the players having much higher salaries. This is likely because each player's usage rate was factored into eABG but not ePER, meaning that players that were efficient over a longer period of time were more likely to shine in eABG. Because of this, we believe that eABG is the most useful metric for measuring player production in relation to their salary.

	player_name	Year
33	Vince Carter	2000.0
339	Todd MacCulloch	2001.0
626	Michael Redd	2002.0
703	Carlos Boozer	2003.0
950	Brian Cardinal	2004.0
1203	Matt Bonner	2005.0
1522	Ryan Gomes	2006.0
1881	Craig Smith	2007.0
2084	Jamario Moon	2008.0
2326	Anthony Morrow	2009.0
2648	Marcus Thornton	2010.0
2744	Landry Fields	2011.0
3057	Isaiah Thomas	2012.0
3347	Brandan Wright	2013.0
3589	Isaiah Thomas	2014.0
3682	Jordan Clarkson	2015.0
4161	Hassan Whiteside	2016.0
4302	Nikola Jokic	2017.0

Figure 8. The highest rated players by ePER each year since 2000.

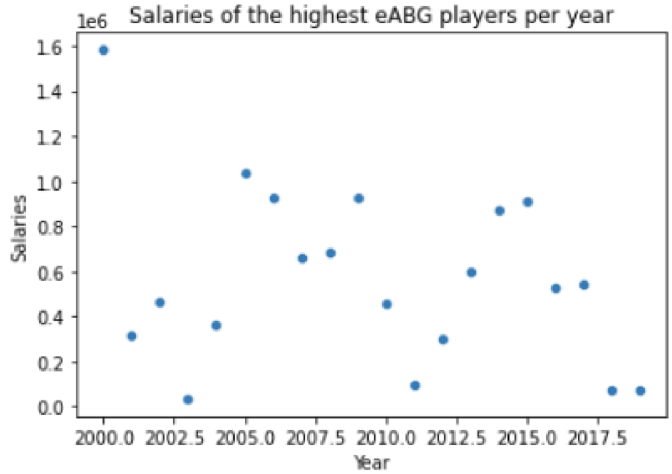

Figure 9. Salary values for each of the top rated eABG players each year since 2000.

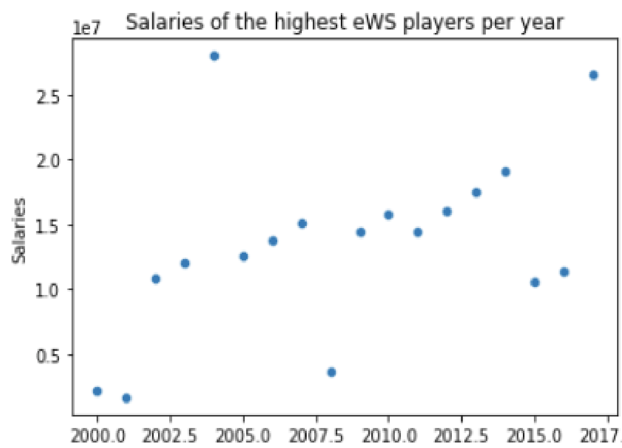

Figure 10. Salary values for each of the top rated eWS players each year since 2000.

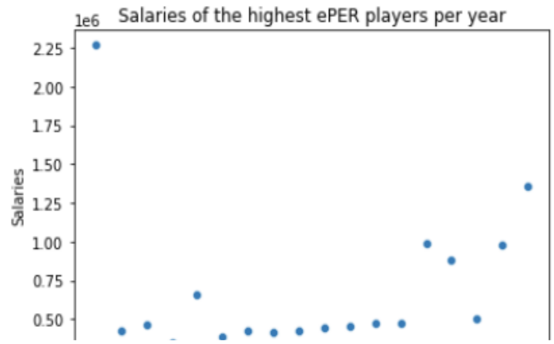

Figure 11. Salary values for each of the top rated ePER players each year since 2000.

Once we decided to move forward with eABG as the best metric for determining player value, we supplemented our analysis of the most productive eABG players by including the player's salary as a proportion of the player's team's total salary. More popular teams tend to be better and have higher total salaries, so we wanted to account for this by including how much the player was paid in comparison to their team's total salary. We determined the players with the highest eABG per year when the percentage of team salary was taken into account (Figure 12). When comparing the two eABG metrics looking at absolute salary and the percentage of team salary, surprisingly 7 players differed between the two different metrics.

We then decided to look at the 5 most productive players in the 2018-2019 season, which was the most recent year in our data set (Figure 13). Interestingly, Duncan Robinson's adjusted eABG score was much, much higher than every other player in 2018-2019. In fact, Duncan Robinson has the highest adj. eABG score in the past 20 years by a decent margin. This makes sense since he is a breakout rookie with a small contract due to being undrafted, but the metric still indicates that he, along with the other 4 high-adj. eABG players, will be worth paying attention to in the coming years.

	player_name	Year
203	Bonzi Wells	2000
328	Eduardo Najera	2001
531	Dan Langhi	2002
935	Devin Brown	2003
1006	Udonis Haslem	2004
1268	Josh Howard	2005
1509	Josh Howard	2006
1860	Ronnie Price	2007
1995	Leon Powe	2008
2352	Matt Barnes	2009
2492	Chase Budinger	2010
2816	Danny Green	2011
2895	Greg Smith	2012
3183	Jae Crowder	2013
3523	Draymond Green	2014
3648	Draymond Green	2015
4084	Jonathon Simmons	2016
4264	Fred VanVleet	2017
4565	Monte Morris	2018
4638	Duncan Robinson	2019

Figure 12. The top-rated adj. eABG players each year.

	Name	Year	Salaries	eABG
4638	Duncan Robinson	2019	77250.0	1701.454959
4626	Donte DiVincenzo	2019	2484360.0	86.147982
4645	Ivica Zubac	2019	1544951.0	72.497353
4627	Brook Lopez	2019	3382000.0	64.342760
4672	Brad Wanamaker	2019	838464.0	51.374295

Figure 13. The 5 highest rate players by adj. eABG in the 2018-2019 season.

Part 5: Limitations and Future Work

One of the limitations we encountered was that NBA salaries differ by several degrees of magnitude. Meanwhile, many of the metrics for production do not vary by much. As such, we had to find a way to make an effective comparison between production metrics that change on a small scale and salaries that change on a very large scale. One of the ways we did this was by taking the log of the salary. However, it is possible this is not the best way to fix the salaries. In the future, we could look into ways to improve how we treat salary in our metrics to allow for more effective comparisons.

Additionally, in each of our metrics we only looked at one player with the highest rating each year in order to make the data easier to interpret. This limited the usefulness of our interpretation of each of the metrics in exchange for ease of use. By looking at more of the highest rated players in each year, we may be able to get a better look at the types of players that are favored in each of the metrics. We may want to look at who ranks at the bottom of each of the metrics in order to see who is not worth their contract.