SPORTS ANALYTICS GROUP BERKELEY

Classification of NBA Salaries through Player Statistics

William Wu, Kevin Feng, Raymond Li, Kunal Sengupta, Austin Cheng
Sports Analytics Group at Berkeley

Abstract

This research creates a classification-based model that predicts the salary amounts given to NBA players in free agency. This was done through collecting player salary data through Spotrac and player statistics through Basketball Reference. Correlation analysis demonstrated that points, turnovers, VORP, and rebounds created a balanced and effective set of variables for unweighted KNN-classification. The former three statistics were standardized by year, while the latter statistic was standardized by year and position due to underlying relationships between those categories. After classification, the model revealed that volume statistics were most indicative of salaries within lower and mid-level players, while high VORP was common among all players making near-max to max salaries. Using test set data, the model was off by 1.21 clusters on average, but many outliers were due to gross over and undervaluation on the part of the franchises. The model was used to predict salaries for incoming free agents in the 2018 class, with reasonable projections that meet qualitative expectations.

1 Motivation

In the NBA, player acquisitions in free agency can often have a critical impact on the fortunes of a franchise. In fact, several dynasties have been built on them - with the Miami Heat of the turn of the decade and the Golden State Warriors of today coming to mind. But perhaps more so than any other form of player acquisition in professional basketball, free agency is a double-edged sword. While the signings of LeBron James and Chris Bosh propelled the Heat to two titles, and that of Kevin Durant and Andre Iguodala solidified the Warriors as a looming dynasty, poor contracts given to players like Joakim Noah, Luol Deng, and Timofey Mozgov have set various franchises back several years, sucking up valuable salary cap space.

In an NBA landscape where teams need several max contract-level players to contend for a championship, filling rosters with high-value players is more important than ever. In order to evaluate players from an analytical perspective, our model sets out to predict the salary value of NBA free agents based on past statistics. The model does not attempt to qualify the inherent value of these statistics (i.e. the expected increase in a team's wins due to the acquisition of a player or the cost of "buying a win" in free agency), but assigns a value to players based off of the contracts that players with similar statistics are given. The intent is for these final results to be taken as an empirical consideration when determining contracts for free agents, a model that could determine the "historical standard" for the salaries of a majority of NBA players.

2 Background

The NBA salary cap is perhaps the most complex out of those implemented by the four major American sporting leagues. While salary caps are designed with parity as their primary goal, the soft cap design of the NBA allows for teams to foster a stronger relationship between their fans and their players. For much of its existence, the cap has remained relatively stable, save for inflation. But over the past five years, the salary cap has exploded due to a rapidly expanding market and lucrative television deals, nearly doubling from 58 million in the 2011-2012 season to 102 million in the current 2017-2018 season. At the beginning of the 2011-2012 season, there was a lockout due to inability to reach an agreement on the salary cap and the division of revenue, as team owners wanted to reduce the percentage of income given to players. Eventually, the owners backed down. Fearing another lockout, the league decided to increase the salary cap and also the share of income given to players. As a result, this has allowed teams to pick up and overpay many free agents they would otherwise be unable to sign.

The NBA salary cap, as it currently stands, contains many variations and exceptions that complicate our analysis. Two of the more pertinent exceptions include the mid-level exception and the Larry Bird exception. The mid-level exception allows for teams to pay a player up to a specified maximum amount without affecting their luxury tax bill. In the 2017-2018 season, the mid-level exception was set at 8.4 million. The Larry Bird exception allows teams to exceed the cap when resigning a qualifying player. There are also several variations of maximum contracts that allow players to be paid different percentages of the cap. These depend on various factors, including the amount of years that the player has been playing for the specific team, and the accolades that the player has accumulated under their tenure. Furthermore, in order to prevent these exceptions from rendering the cap moot, a luxury tax penalizes teams for exceeding the salary cap by large amounts of salary.

The salary cap's many nuances presented a significant challenge throughout our analysis. The large increase in the cap sizing made free agent salaries hard to scale, as the willingness of general managers to overpay increased greatly, and salaries quickly began taking more and more of the cap. This forced us to limit our modeling to contracts signed from the 2014-15 offseason onwards, mainly because contracts sizes and lengths given to players had increased greatly from

2011-12. We also needed to scale and analyze salaries by their share of the year's salary cap, to better standardize salaries throughout the cap's rapid growth. Meanwhile, the various exceptions and salary standards the CBA make salaries in the NBA come in more of a tiered approach. For example, the minimum salary, maximum salary, and mid-level exception are all determined by cap sizing. As a result, we take a more discrete approach in our modeling, grouping salaries into clusters rather than analyzing them through a continuous scale.

3 Data Sources

We used Spotrac's online database to both generate yearly lists of NBA free agents and collect data pertaining to contract signings within each year. For this study, we focused on average yearly salary, ignoring the typical back-end weighing of NBA salaries due to the contracts being fully guaranteed. We standardized player salaries and projections based off of the cap numbers and projections listed in RealGM.

For player statistics, we used seasonal data collected by Basketball Reference, evaluating all metrics listed in the Per-Game, Per 36 Minutes, Per 100 Possessions, and Advanced data tables under each relevant season.

We were able to collect data on nearly 500 contracts signed in between 2014 and 2017 with which to build our model. We filtered this set to remove all players who had signed multiple contracts within a single season, as well as any contract under $300,000 dollars, which were mainly composed of 10-day and two-way contracts. We excluded players within the former category entirely because all of their contracts were almost always part of the latter. Since our study pertains primarily to the initial free agent valuation of players, our focus was towards the salary within a player's single contract signing, even if the combined contract amount would exceed the $300,000 cutoff. After filtering, 455 contracts remained eligible to be used for the model, a sufficient sample size that contained players earning salaries through all ends of the spectrum.

Furthermore, we also reference Spotrac's tracker of upcoming 2018 free agents, a list that does not include players with pending player or team-options. Thus, we do not make predictions on salaries for players such as LeBron James, who many believe will enter free agency later this year.

4 Methodology

4.1 Variable Selection

Initially, the data from Spotrac and Basketball Reference was analyzed in conjunction in order to observe the individual relationships between the various statistics and player salary. We looked into player data in the 2016-17 Season and free agent signings in the 2017 offseason to choose our variables with which to build our model. This was done through the following correlation table below:

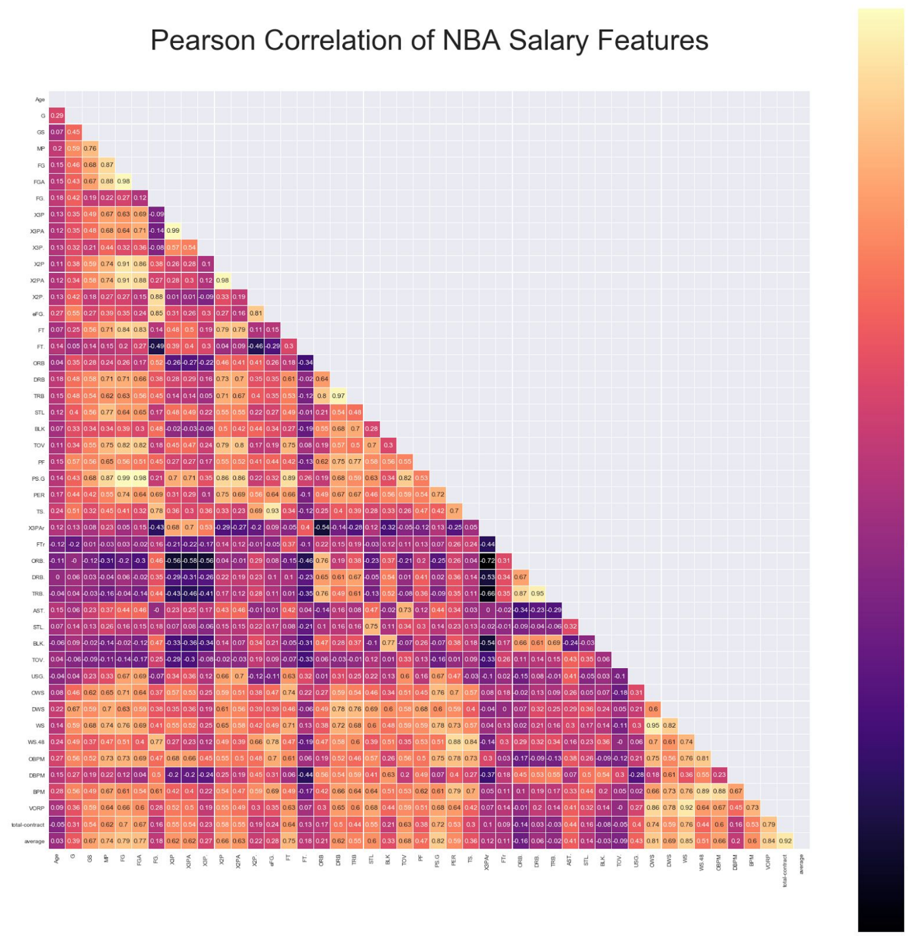


Figure 4.1.1 Initial correlation table used for variable selection, with analysis of nearly 40 variables

Using this table, we were able to eliminate stats such as age, games played, field goal percentage, and most defensive measures, due to weaker correlations with salary. Afterwards, we took a closer look onto 12 variables, this time looking to eliminate variables that correlated too well with each other to avoid building a model on different statistics that essentially told the same thing.

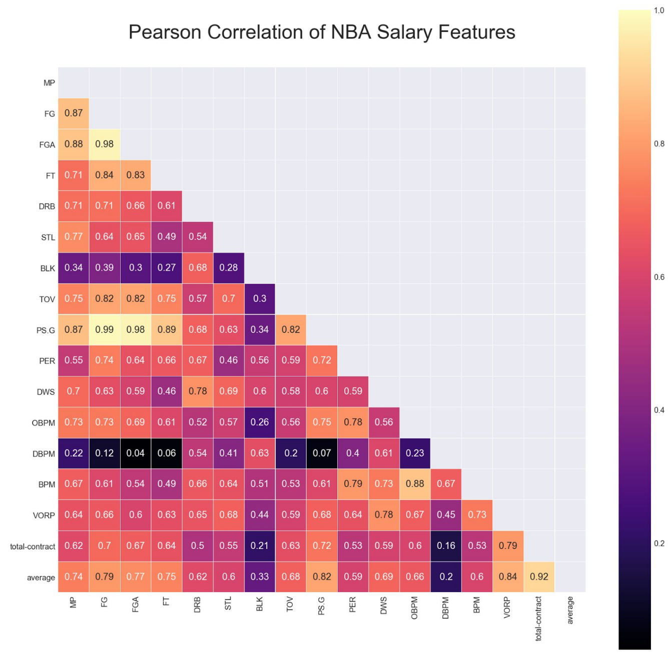


Figure 4.1.2 Reduced correlation table, used to reduce statistical noise due to overly correlated dimensions

From this, we settled on four variables: defensive rebounds, turnovers, and points (all per-game), as well as VORP (value over replacement player). All of these correlated well with salary, with the holistic nature of VORP, which correlated strongly with many offensive and defensive statistics, ensuring that the model covered a good amount of breadth. While turnovers did correlate strongly with points, we felt that there was enough of a qualitative difference to allow it into the model.

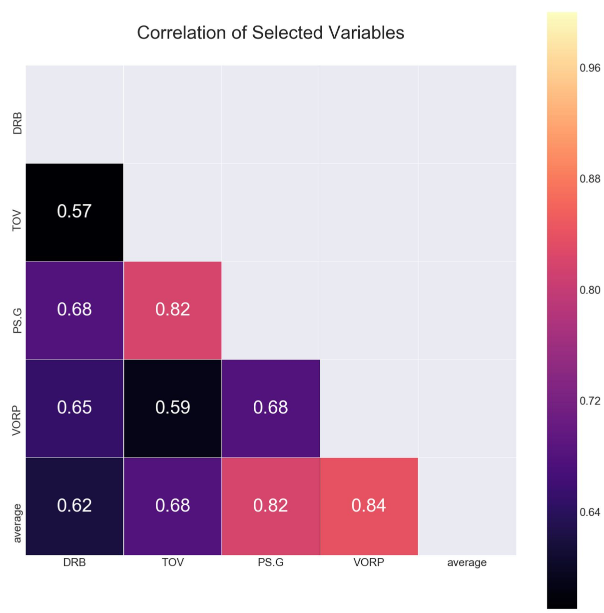


Figure 4.1.3 Final correlation table of selected variables

4.2 Standardization

Afterwards, each of the four variables were standardized using the respective means and standard deviations among the free agent class of each same year. Players were standardized as an entire free agent class for PPG, TOV, and VORP, but each had their DRB stats standardized according to their position group. The histograms below display why this was necessary, as the distribution of DRB among players in 2017 differed systematically by position. To deal with this, Point and Shooting Guards were grouped, as well as were Small and Power Forwards, with Centers being left by themselves.

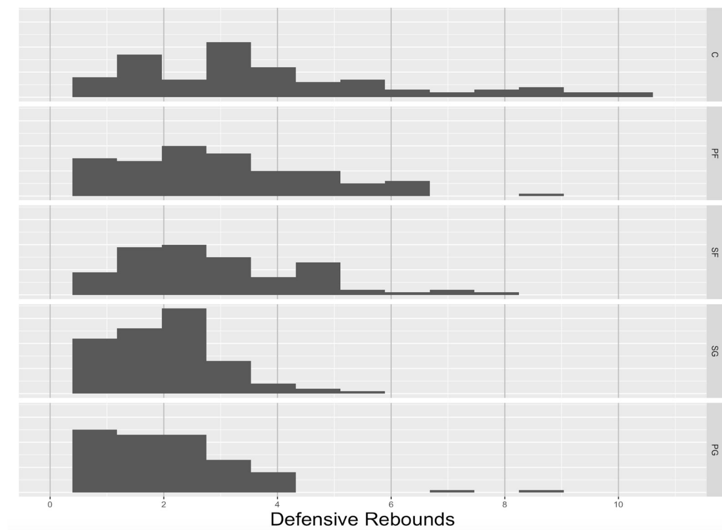


Figure 4.2.1 Distribution of DRB by position

Because of salary inflation throughout recent years, as well as wide variability among salaries themselves, we took several steps in simplifying and standardizing our salary data. Salaries were standardized into a percent-cap value, the percentage of cap space (of the year of the signing) taken by the average salary size of the contract. This allowed us to better take into account salaries relative to the offseason of signing, which is especially important considering the large jump in salary cap space from 2014 to 2018.

Year	Cap Space
2018-19	$101,000,000 (Projected)
2017-18	$99,093,000
2016-17	$94,143,000
2015-16	$70,000,000
2014-15	$63,065,000

Table 1 Cap space over 2014-18, with next year's projection

Afterwards, the salaries were grouped into clusters that were formed to encapsulate several of the salary standards within the NBA. The clusters' bounds are based off of the distribution of percent-cap shown below:

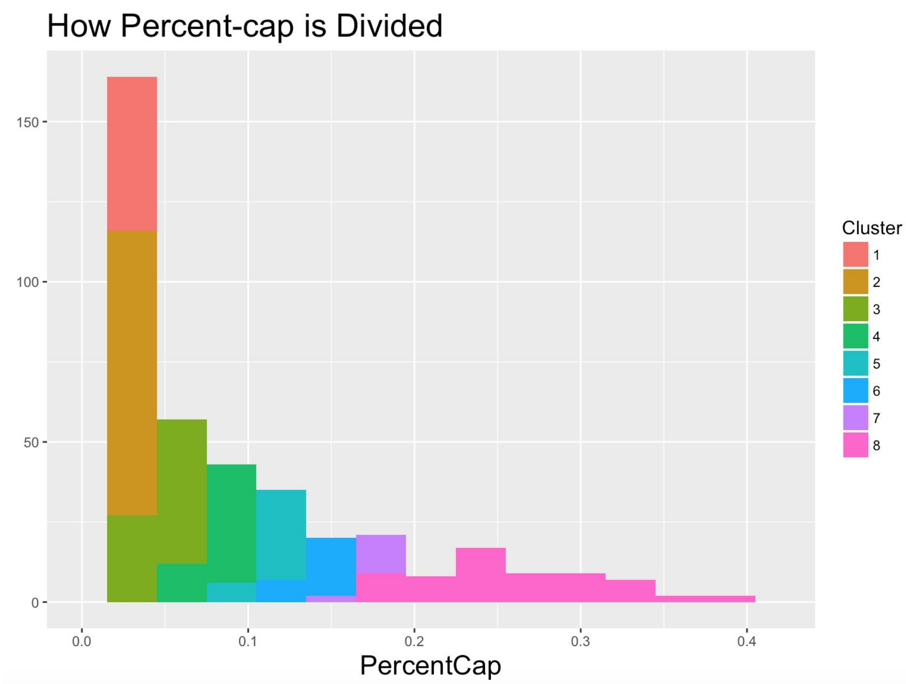


Figure 4.2.2 Percent-cap distribution, with the cluster division highlighted

Percent Cap	**Cluster**
< 2% (Min)	1
2-4% (Vet Min)	2
4-7% (MLE, Bi-annual)	3
7-10%	4
10-13%	5
13-16%	6
16-19%	7
> 19% (Maxes)	8

Table 2 Clustering bounds used to group data, with various exceptions listed

The bounds are narrower at lower percentages, distinguishing between several different contract exceptions and minimum salary rates. They are wider among higher salaries, to account for their higher variability. All near-max and max salaries are included in a single cluster, to avoid the trouble of needing to distinguish between the various forms of max contracts, which are decided by a variety of rules involving team affiliation and player accolades. This lead to the final distribution shown in Figure 4.2.3, which maintains part of the shape but is slightly more uniformed.

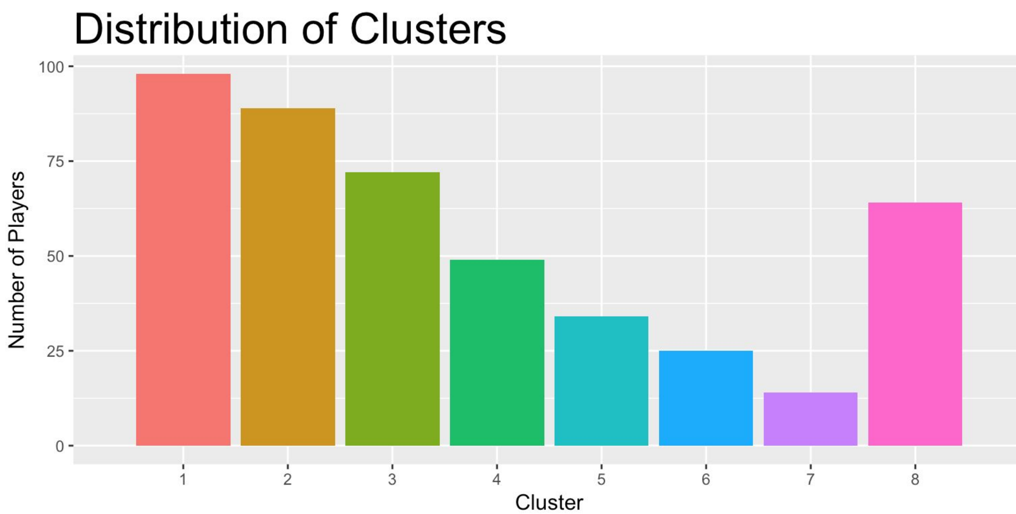


Figure 4.2.3 Distribution of clusters in the data used to train our model

4.3 Classification

After the standardization and clustering process, the dataset was used to train a K-Nearest-Neighbors classification model. The five rows of the source dataset are shown below as an example.

Name	DRB	TOV	PPG	VORP	Salary	Percent Cap	Cluster
Leandro Barbosa	-0.229616803	-0.211584266	-0.100782033	-0.619041863	1448490	0.022968207	2
Eric Bledsoe	2.139833181	2.631595332	1.640816521	0.882609762	14000000	0.221993182	8
Avery Bradley	1.097275189	0.53011476	1.162730643	-0.619041863	8000000	0.126853247	5
Rasual Butler	-1.082618797	-1.200516299	-0.920357823	-0.393794119	1448490	0.022968207	2
Vince Carter	0.81294119	0.159265247	0.650495774	0.732444599	4088019	0.06482231	3

Table 3 First five rows of the dataset used to train the KNN model

We used an 80-20 training-test split, done through simple random selection. After training, the model was used to predict the salaries of upcoming NBA free agents in 2018, based off of player statistics as of January 11, 2018. To address sample size concerns for player statistics, the model was only used to predict salaries for players who had played at least 5 minutes per game over at least 5 regular season games. Classification was performed on the test set using the 17 nearest-neighbors in unweighted Euclidean distance between the four variables. This allowed us to make the most conservative estimate at a reasonable error size.

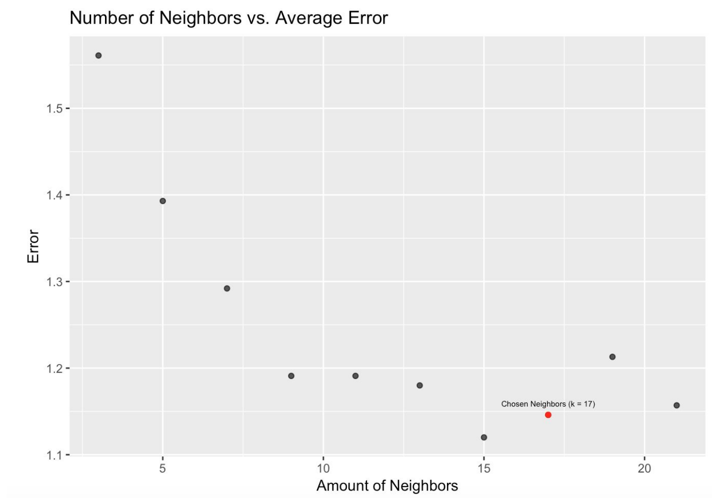


Figure 4.3.1 Scatterplot showing the average error in the test set vs. the amount of neighbors used for classification

4.4 Results

Player	Pos	Age	Tm	Cluster	Predicted Salary		Player	Pos	Age	Tm	Cluster	Predicted Salary
Arron Afflalo	SG	32	ORL	1	< $2,020,000		Davis Bertans	PF	25	SAS	2	$2,020,000 - $4,040,000
Will Barton	SG	27	DEN	2	$2,020,000 - $4,040,000		Nemanja Bjelica	PF	29	MIN	1	< $2,020,000
Marco Belinelli	SG	31	ATL	2	$2,020,000 - $4,040,000		Corey Brewer	SF	31	LAL	1	< $2,020,000
Avery Bradley	SG	27	DET	5	$10,100,000 - $13,130,000		Vince Carter	SF	41	SAC	1	< $2,020,000
Aaron Brooks	PG	33	MIN	1	< $2,020,000		Omri Casspi	SF	29	GSW	3	$4,040,000 - $7,070,000
Jose Calderon	PG	36	CLE	1	< $2,020,000		Nick Collison	PF	37	OKC	1	< $2,020,000
Kentavious Caldwell-Pope	SG	24	LAL	3	$4,040,000 - $7,070,000		Dante Cunningham	SF	30	NOP	1	< $2,020,000
Michael Carter-Williams	PG	26	CHO	2	$2,020,000 - $4,040,000		James Ennis	SF	27	ME M	4	$7,070,000 - $10,100,000
Mario Chalmers	PG	31	MEM	3	$4,040,000 - $7,070,000		Aaron Gordon	PF	22	ORL	8	> $19,190,000
Ian Clark	SG	26	NOP	1	< $2,020,000		Jerami Grant	PF	23	OKC	3	$4,040,000 - $7,070,000
Pat Connaughton	SG	25	POR	1	< $2,020,000		Jeff Green	PF	31	CLE	2	$2,020,000 - $4,040,000
Wayne Ellington	SG	30	MIA	4	$7,070,000 - $10,100,000		Mario Hezonja	SF	22	ORL	2	$2,020,000 - $4,040,000
Tyreke Evans	SG	28	MEM	8	> $19,190,000		Josh Huestis	PF	26	OKC	1	< $2,020,000
Raymond Felton	PG	33	OKC	3	$4,040,000 - $7,070,000		Ersan Ilyasova	PF	30	ATL	3	$4,040,000 - $7,070,000
Yogi Ferrell	SG	24	DAL	5	$10,100,000 - $13,130,000		Richard Jefferson	SF	37	DEN	1	< $2,020,000
Bryn Forbes	SG	24	SAS	1	< $2,020,000		Joe Johnson	SF	36	UTA	3	$4,040,000 - $7,070,000
Tim Frazier	PG	27	WAS	1	< $2,020,000		Chris McCullough	PF	22	WAS	1	< $2,020,000
Treveon Graham	SG	24	CHO	2	$2,020,000 - $4,040,000		Doug McDermott	SF	26	NYK	2	$2,020,000 - $4,040,000
Gerald Green	SG	32	HOU	3	$4,040,000 - $7,070,000		Jordan Mickey	PF	23	MIA	3	$4,040,000 - $7,070,000
Devin Harris	PG	34	DAL	4	$7,070,000 - $10,100,000		Johnny O'Bryant	PF	24	CHO	1	< $2,020,000
Joe Harris	SG	26	BRK	5	$10,100,000 - $13,130,000		Quincy Pondexter	SF	29	CHI	1	< $2,020,000

Player	Pos	Age	Tm	Cluster	Predicted Salary		Player	Pos	Age	Tm	Cluster	Predicted Salary
Rodney Hood	SG	25	UTA	1	< $2,020,000		Anthony Tolliver	PF	32	DET	2	$2,020,000 - $4,040,000
Jarrett Jack	PG	34	NYK	2	$2,020,000 - $4,040,000		Noah Vonleh	PF	22	POR	1	< $2,020,000
Shane Larkin	PG	25	BOS	1	< $2,020,000		Okaro White	PF	25	MIA	1	< $2,020,000
Patrick McCaw	SG	22	GSW	1	< $2,020,000		Brandan Wright	PF	30	MEM	2	$2,020,000 - $4,040,000
Shabazz Napier	PG	26	POR	5	$10,100,000 - $13,130,000		Trevor Booker	PF	30	TOT	2	$2,020,000 - $4,040,000
Jameer Nelson	PG	35	NOP	1	< $2,020,000		Aron Baynes	C	31	BOS	3	$4,040,000 - $7,070,000
Raul Neto	PG	25	UTA	1	< $2,020,000		Tarik Black	C	26	HOU	1	< $2,020,000
David Nwaba	SG	25	CHI	5	$10,100,000 - $13,130,000		Clint Capela	C	23	HOU	2	$2,020,000 - $4,040,000
Tony Parker	PG	35	SAS	3	$4,040,000 - $7,070,000		DeMarcus Cousins	C	27	NOP	8	> $19,190,000
Chris Paul	PG	32	HOU	8	> $19,190,000		Ed Davis	C	28	POR	2	$2,020,000 - $4,040,000
Elfrid Payton	PG	23	ORL	1	< $2,020,000		Derrick Favors	C	26	UTA	5	$10,100,000 - $13,130,000
J.J. Redick	SG	33	PHI	8	> $19,190,000		Channing Frye	C	34	CLE	1	< $2,020,000
Rajon Rondo	PG	31	NOP	2	$2,020,000 - $4,040,000		Montrezl Harrell	C	24	LAC	2	$2,020,000 - $4,040,000
Derrick Rose	PG	29	CLE	2	$2,020,000 - $4,040,000		Udonis Haslem	C	37	MIA	1	< $2,020,000
Ramon Sessions	PG	31	NYK	2	$2,020,000 - $4,040,000		Amir Johnson	C	30	PHI	3	$4,040,000 - $7,070,000
Marcus Smart	SG	23	BOS	2	$2,020,000 - $4,040,000		Alex Len	C	24	PHO	1	< $2,020,000
Julyan Stone	SG	29	CHO	1	< $2,020,000		Kevon Looney	C	21	GSW	1	< $2,020,000
Jason Terry	SG	40	MIL	1	< $2,020,000		Brook Lopez	C	29	LAL	2	$2,020,000 - $4,040,000
Fred VanVleet	PG	23	TOR	4	$7,070,000 - $10,100,000		JaVale McGee	C	30	GSW	2	$2,020,000 - $4,040,000
Rashad Vaughn	SG	21	MIL	1	< $2,020,000		Salah Mejri	C	31	DAL	4	$7,070,000 - $10,100,000
Dwyane Wade	SG	36	CLE	1	< $2,020,000		Nerlens Noel	C	23	DAL	2	$2,020,000 - $4,040,000
Nick Young	SG	32	GSW	1	< $2,020,000		Lucas Nogueira	C	25	TOR	1	< $2,020,000

Player	Pos	Age	Tm	Cluster	Predicted Salary		Player	Pos	Age	Tm	Cluster	Predicted Salary
Isaiah Canaan	PG-SG	26	TOT	2	$2,020,000 - $4,040,000		Jusuf Nurkic	C	23	POR	2	$2,020,000 - $4,040,000
Sean Kilpatrick	SG	28	TOT	1	< $2,020,000		Zaza Pachulia	C	33	GSW	3	$4,040,000 - $7,070,000
Nik Stauskas	SG	24	TOT	1	< $2,020,000		Julius Randle	C	23	LAL	1	< $2,020,000
Quincy Acy	PF	27	BRK	3	$4,040,000 - $7,070,000		Willie Reed	C	27	LAC	1	< $2,020,000
Tony Allen	SF	36	NOP	1	< $2,020,000		Marreese Speights	C	30	ORL	1	< $2,020,000
Kyle Anderson	SF	24	SAS	2	$2,020,000 - $4,040,000		David West	C	37	GSW	4	$7,070,000 - $10,100,000
Trevor Ariza	SF	32	HOU	4	$7,070,000 - $10,100,000		Greg Monroe	C	27	TOT	1	< $2,020,000
Luke Babbitt	SF	28	ATL	1	< $2,020,000		Jahlil Okafor	C	22	TOT	1	< $2,020,000
Michael Beasley	PF	29	NYK	2	$2,020,000 - $4,040,000							

Table 4 Predictions of salaries for incoming free agents in 2018 based off player data as of January 11, 2018.

5 Error Analysis

An L1-norm error function (eq. 1), also known as least absolute errors, was used to tune and evaluate the performance of the model. This function was chosen over other popular error functions such as the L2-norm due to its interpretability and to avoid over or under-punishing outliers. The average error of the test set was a cluster size of 1.21.

$$S = \sum_{i=1}^{n} |y_i - f(x_i)| \tag{1}$$

In analyzing the biggest factors in this error, we explored the average statistics per actual and predicted salary cluster, as shown in Table 5 below. While PPG and DRB follow similar distributions for actual and predicted clusters, the distributions for TOV and VORP vary substantially. This suggests that PPG and DRB are the two most predictive statistics of salary. This makes sense intuitively as NBA contracts have largely been based on raw statistics in the past, as the focus on advanced analytics is a new trend. Also noted is that there were no players predicted in cluster 7 - this is to be expected given the high k-neighbors value relative to the amount of total players in cluster 7. This also makes sense intuitively, as there are very few players who receive just below max contracts.

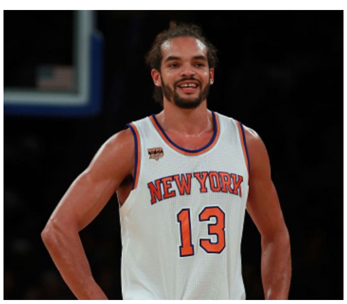
	Average Z-Score by Cluster (Actual | Predicted)							
Cluster	PPG		DRB		TOV		VORP	
1	-0.539	-0.665	-0.574	-0.695	-0.531	-0.668	-0.443	-0.481
2	-0.300	-0.017	-0.316	-0.072	-0.229	0.078	-0.307	-0.337
3	-0.043	0.016	0.064	0.222	-0.084	-0.018	-0.272	-0.223
4	0.140	0.063	0.238	0.210	-0.047	-0.120	0.243	0.486
5	0.361	0.779	0.402	0.403	0.170	0.234	0.166	-0.069
6	0.680	0.699	0.798	1.118	0.660	0.325	0.513	-0.298
7	0.471	N/A	0.764	N/A	0.386	N/A	0.855	N/A
8	1.502	1.441	1.316	1.324	1.088	1.136	1.512	1.506

Table 5 Average statistics per actual and predicted cluster

While this error function provided a sound mathematical basis for optimizing the model and evaluating its performance relative to other models, it was impossible to evaluate its performance relative to the true value of each player. For example, the model predicted Joakim Noah's (New York Knicks) salary to be within cluster 1, while his actual salary is within cluster 8, resulting in the maximum L1 error of 7. Although this is empirically a highly erroneous prediction, it can be argued that Joakim Noah's true value is much closer to his predicted salary than his actual salary. Such observations often rely on expert opinions and off-the-statsheet analyses to validate. The top 3 erroneous predictions will be analyzed below as an observational evaluation of model performance.

Player	Actual Cluster	Predicted Cluster
Joakim Noah	8	1
Reggie Jackson	8	1
Bismack Biyombo	8	3

Table 6 Largest outliers



A former All-Star, Joakim Noah received a big contract from the New York Knicks, who were hoping for Noah to regain his form after a down year due to injury. Unfortunately, Noah has been a shell of his former self ever since and has recorded career-lows in almost all statistical categories. Having played less than 40 minutes thus far in the 2017-2018 season, Joakim Noah is by almost all measures one of the most overpaid players in the league.

Following a knee injury in 2016, Reggie Jackson has yet to display the same explosiveness as when he signed his big contract. A below average shooter with a VORP of almost 0, it is evident why Jackson's predicted salary is so low. However, very few experts would consider him a minimum-level player, and his true value is likely between his predicted and actual salaries.

Bismack Biyombo signed his four-year $72M contract in 2016 after a strong playoff run with the Raptors the previous season. However, that run has proven to be an outlier, as Biyombo has not showed the same level of promise during his time with the Magic. Nothing more than a decent defensive player, Biyombo's true value is likely much closer to his predicted salary than his actual salary.

It is to be expected that the model is more prone to predicting that players are being overpaid, as there are generally more overpaid than underpaid players due to the nature of contract negotiations. History has shown that it is much more likely for players to get a big contract and fail to live up to expectations due to injuries or extraneous factors than it is for small contract players to suddenly break out and greatly outperform his contract. This is especially the case when excluding rookie contracts from the analysis.

While the true value of a player is, by nature, impossible to boil down to a science, this observational evaluation suggests that our model is robust in predicting true player values rather than simply predicting current player contracts. To be tuned to a certain person (or team's) preferences, a similar observational evaluation would be necessary in determining whether the model follows his or her perception of true value.

6 Analysis

6.1 Model Trends

	Average Z-Score per Cluster			
Cluster	PPGS	DRB	TOV	VORP
1	-0.665	-0.695	-0.668	-0.481
2	-0.017	-0.072	0.078	-0.337
3	0.016	0.222	-0.018	-0.223
4	0.063	0.21	-0.12	0.486
5	0.779	0.403	0.234	-0.069
6	0.699	1.118	0.325	-0.298
7	N/A	N/A	N/A	N/A
8	1.441	1.324	1.136	1.506

Table 7 Typical player within each cluster, by Z-Score

Our model tends to weigh high volume over high efficiency in its predictions. Out of the four statistics from which the model is based, VORP has the weakest relationship with the dependent variable. Although top-paid players tend to have high VORPs, the relationship for others is noisy, at best. For example, there several free agents were predicted to have salaries within the 6th cluster with sub-par VORPs, as they made up for it in the other categories.

Among those volume statistics, points per game had the strongest linear relationship between through clusters, followed by defensive rebounds and turnovers. The relationship in the latter is particularly interesting, as it jumps from .325 among players in Cluster 6 to 1.136 among those within Cluster 8. Because this drastic increase is much greater relative to the increase that we see in scoring, we can infer that players who earn near-max to max contracts tend to be those who are responsible for a large share of their team's offensive possessions, regardless of whether or not they are relied upon to use those possessions to score consistently.

6.2 2018 Free Agency Demographics

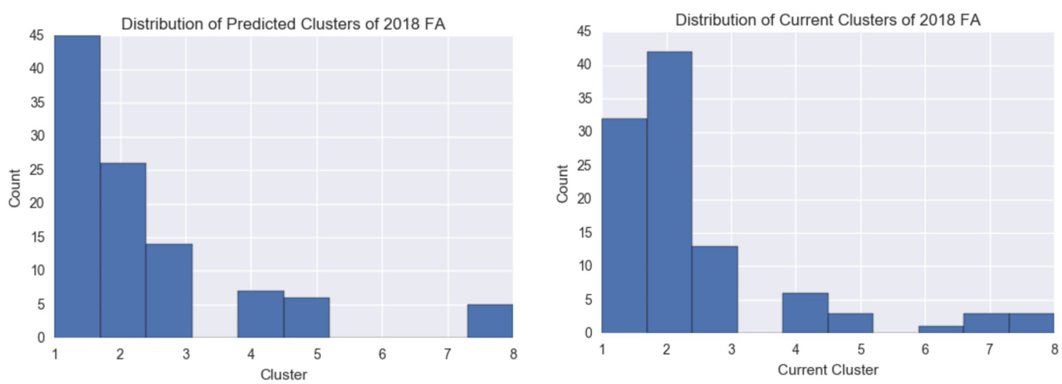


Figure 6.2.1 Distribution of current and predicted salaries of incoming 2018 free agents

Similar to most job salaries, the distribution of our predicted free agent salaries is right skewed as the majority of players are in lower clusters. This is similar to the distribution of the current cluster salaries of the players.

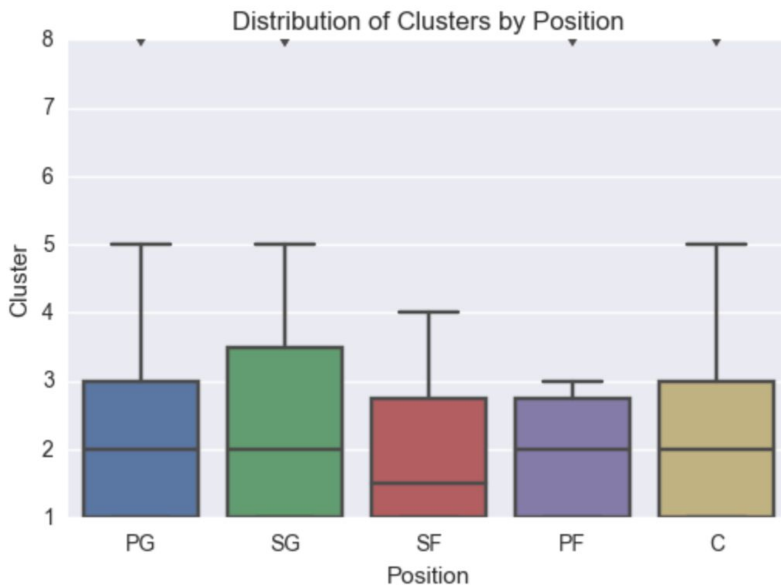


Figure 6.2.2 Distribution of predicted salaries by player position

By looking at our results positionally, we can see that the upcoming free agent class is particularly weak at small forward, with a lower median and quartiles. The position also lacks star talent, with no free agent small forwards with a salary predicted within Cluster 8. In general, though, there seems to be no definitive relationship between position and projected compensation.

Figure 6.2.3 Distribution of the differences between predicted and current salaries by franchise

Although the Lakers are currently suffering by overpaying their players, they look to benefit the most this offseason as many of their upcoming free agents such as Kentavious Caldwell-Pope and Brook Lopez are on overpaid expiring contracts. The Memphis Grizzlies, on the other hand, will be hurt as upcoming free agents Tyreke Evans and James Ennis may be seeking bigger deals.

6.3 Salary Movement

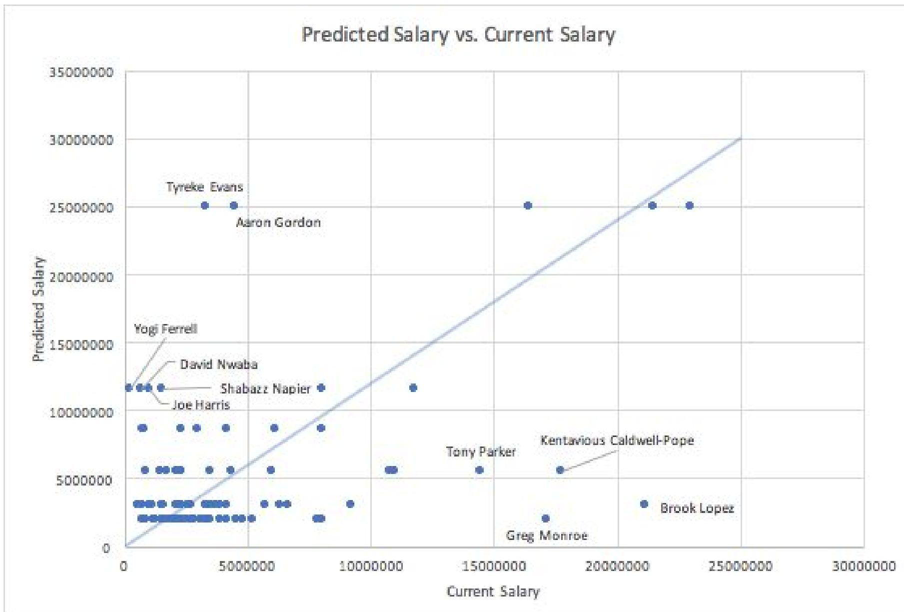
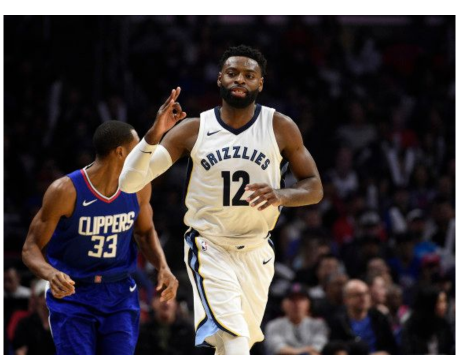


Figure 6.3.1 Predicted vs. current salary, using midpoints of bins

By taking the midpoint of the salary bins for each cluster, and taking the 8th cluster to signal a maximum salary, we visualized the results into scatterplot form. Through this, we can get a rough picture of the largest upcoming pay raises and pay cuts.

Notable Players Looking for a Pay Increase



The 2010 Rookie of the Year, Tyreke Evans' career stagnated until signing with the Memphis Grizzlies on a 1 year / $3.3 million contract this past offseason. Filling in for the injured Mike Conley, Evans has averaged 19.4 points, 5.1 rebounds, 5.1 assists, on 45.3/38.8/79.7 shooting splits. These stats project him in line with a top of the line salary of more than $19 million. Although he is unlikely to receive a contract of such size due to common belief that his stats are inflated by playing on a terrible Grizzlies team, Evans has nonetheless displayed a terrific year similar to his rookie year form, which would put him in line for a major pay increase.

Aaron Gordon is on the last year of his rookie deal, and will demand a heavy contract through his play this season. Already a high-flier, Gordon has shown steady improvement with his jumper this season. He is averaging 18.3 points, 8.3 rebounds, 2.1 assists on 44.7/35.1/73.5 shooting splits. At only 22 years old, Gordon's potential and athleticism project him for a max contract.

As a 6' undersized guard, Yogi Ferrell went undrafted in the 2016 NBA draft. After signing a 10-day contract with the Mavericks, Ferrell's play earned him a two-year contract that pays him $1.3 million in the 2017-18 season. This year, Ferrell has played great off the Dallas bench, averaging 10.4 points, 3.5 rebounds, 2.3 assists, to go along with 40.9% three point shooting. This puts him in line for an eight-figure salary, which Ferrell will look to secure as a 2018 restricted free agent.

Once thought of as a bust, Shabazz Napier has revitalized his career on the Portland Trail Blazers. Currently averaging 9.8 points, 2.4 rebounds, 2.3 points, with 42.2% three point shooting, Napier has been one of the best backup point guards in the league. As such, he is projected for a raise from $2.4 million to around $10 million.

Notable Players Looking at Major Pay Cuts

Brook Lopez, a talented big on Brooklyn, has drastically declined in the 2017-18 season. As a center, he is shooting close to 40% from the field which is quite low compared to his career average which is 50.2%. Not gelling well with the Lakers offense, he has lost his starting position and is now coming off the bench, seeing less than 20 minutes a game. As he was accustomed to being the centerpiece of an offense in Brooklyn, he now has to change his playstyle to fit the Lakers. He will most likely see a large pay cut next season as he is severely overpaid this year with a 22 million dollar contract.

The 35 year old veteran, Tony Parker has definitely hit the latter end of his career, and is continuing to show his age the past couple of years. Being a key member of the Spurs for so long, he is still making around $15 million even though his value has steadily declined. Contributing only

8.6 points with 4.5 assists, he is only a minor part of the Spurs offense. Defensively, he has worn down and is not able to keep up with the fast paced guards in today's game.

Although Greg Monroe had a fantastic start to his career, he has been declining in his prime years. In the new era of big men who can shoot and dribble, Monroe's outdated playing style has left him with mediocre stats as he is scoring only 11 points per game while grabbing 8.5 boards a game. He should see a salary decrease from $17 million a year to less than $5 million.

7 Conclusion

In a landscape where obtaining high-value contracts is critical in the success of a franchise, the rise of advanced analytics has enabled more accurate empirical valuations of players. Our model has proven to be a robust predictor of player value from both an empirical and observational standpoint. Further research is still necessary to prevent the model from overvaluing high volume, "empty stats" players - an issue that GMs still struggle with today. In analyzing the results that of our model, we came to the conclusion that our model values high volume scorers over more efficient role players. There is a consistent correlation between PPG, DRB and TOV and predicted salary throughout the entire NBA pay scale. While the adage that sports are not played on spreadsheets certainly holds true, and the "eye-test" is still a key component of evaluating players, advanced analytics using machine learning is the next step in more accurately determining a player's exact value.

8 Afterword

8.1 GitHub Repository

https://github.com/wwu2020/KNN-NBA-FA

8.2 References

[1] Spotrac Free Agents and Salary Database. http://www.spotrac.com/nba/free-agents/
[2] NBA Salary Cap History and Projections. https://basketball.realgm.com/nba/info/salary_cap
[3] Basketball Reference. https://www.basketball-reference.com/
[4] CBA Salary Cap FAQ and Information. http://www.cbafaq.com/salarycap.htm

8.3 Images

[1] Joakim Noah
http://www.espn.com/nba/player/_/id/3224/joakim-noah
[2] Reggie Jackson
http://www.mlive.com/pistons/index.ssf/2017/03/as_playoff_hopes_dim_pistons_p.html
[3] Bismack Biyombo
https://orlandomagicdaily.com/2017/01/14/orlando-magics-biggest-first-half-disappointment/
[4] Tyreke Evans
https://fivethirtyeight.com/features/tyreke-evans-is-back/
[5] Aaron Gordon
http://www.hiphoptitans.com/slam-dunk-contest-aaron-gordon-insane-dunk-nba-all-star-game-2016/
[6] Yogi Ferrell
https://www.si.com/nba/2017/02/06/yogi-ferrell-10-day-contract-mavericks-nets-hoosiers
[7] Greg Monroe
https://cdn.nba.net/nba-drupal-prod/2017-11/_monroe_iso_aotn.jpg
[8] Tony Parker
http://www.azquotes.com/public/pictures/authors/99/1f/991f321f9d81ea2f530c99a5812a2ac6/54b85d339e7d6_tony_parker.jpg
[9] Brook Lopez
https://www.silverscreenandroll.com/2017/12/1/16695194/la-lakers-brook-lopez-shooting-slump-analysis
[10] Shabazz Napier
http://a1.espncdn.com/combiner/i?img=%2Fphoto%2F2016%2F0930%2Fr133547_1296x729_16-9.jpg

8.4 Technology

[1] NumPy, SciPy, http://www.numpy.org/ https://www.scipy.org/
[2] scikit-learn http://scikit-learn.org/
[3] R, R-Studio https://www.r-project.org/ https://www.rstudio.com/
[4] ggplot2 http://ggplot2.tidyverse.org/index.html