Predicting NBA Success

General Assembly Data Science Immersive Capstone Project
Matt Lee

Can we accurately predict the NBA potential of college basketball players so that NBA teams can make the best use of their draft pick?

NBA Draft

- **30 teams**
- **2 rounds**
- **60 picks**

What's the problem?

1st Round, 1st Pick
From top-left:
Anthony Bennett (2013), Kwame Brown (2001),
Greg Oden (2007), Andrea Bargnani (2006)

2nd Round
From top-left:

Isaiah Thomas (2011, **60th**), Marc Gasol (2007, **48th**),
Draymond Green (2012, **35th**), Manu Ginobili (1999, **57th**)

Significance of the draft

Once a year opportunity to add value

- Improve team roster

- Make history

- Boost ticket and merchandise sales

- Raise awareness to local community

- Aid and boost local community and economy

Approach

- Collect pre-NBA data that can capture as many features of basketball players as possible
  - College basketball career average
  - Physical features
  - Demographic features

- Identify target metrics that can accurately represent the skill of an NBA player in his first five years of playing

- Generate an ordered list of players ranked by the predicted metric for team(s)

Limitations

- Data acquisitions
  - Availability
    - Structured demographic data
    - High school/International players
  - Compatibility
    - No primary keys
    - Trades/transfers
  - Timing
    - Different eras
    - Measurement recording period
    - Deferred debuts/mid-career breaks

- Domain knowledge on target metric
  - No perfect statistic exists
  - Defense is hard to measure

Target Metric

- Goal: Produced an ordered list, so teams know what they are getting in return

- To allow for flexibility, generate lists ranked in terms of:
  - **Overall Skill**
  - **Offense**
  - **Defense**

- Advanced basketball metrics (PER, Win Shares, Box +/-, VORP, etc.)

- To account for weaknesses, utilized dimensionality reduction (PCA) and unity-based normalization to generate one single metric for each list
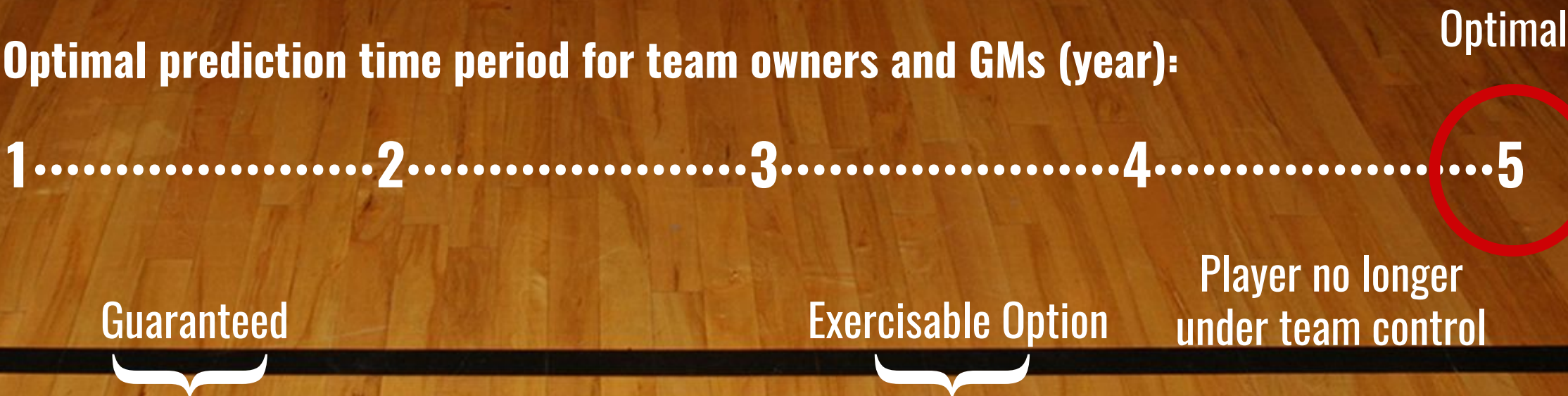
Target Metric - Timing

Rookie Contract

* First round picks can sign for as much as 120% and as little as 80% of the rookie scale.

Pick	1st Year Salary	2nd Year Salary	3rd Year Option Salary	4th Year Option: Percentage Increased Over 3rd Year Salary	Qualifying Offer: Percentage Increase Over 4th Year Salary
1	$4,919,300	$5,140,700	$5,362,100	26.1%	30.0%
2	$4,401,400	$4,599,500	$4,797,600	26.2%	30.5%
3	$3,952,500	$4,130,400	$4,308,300	26.4%	31.2%

Optimal prediction time period for team owners and GMs (year):

Optimal

1 .... 2 .... 3 .... 4 5

Guaranteed

Exercisable Option

Player no longer under team control

Modelling

Regression Problem

- Initially, ran trial models on all input variables

  - Linear Regressions, KNeighbors, Decision Tree, Ensemble and Boosted Regressors

- Data fitted better on linear models

  - Data transformation to reduce skewness and make patterns more linear

  - Utilized SelectKBest and models with regularization to identify key features

  - Train test split and K-Fold cross-validation

- Target metric combinations with unity-based normalization had better results than PCA
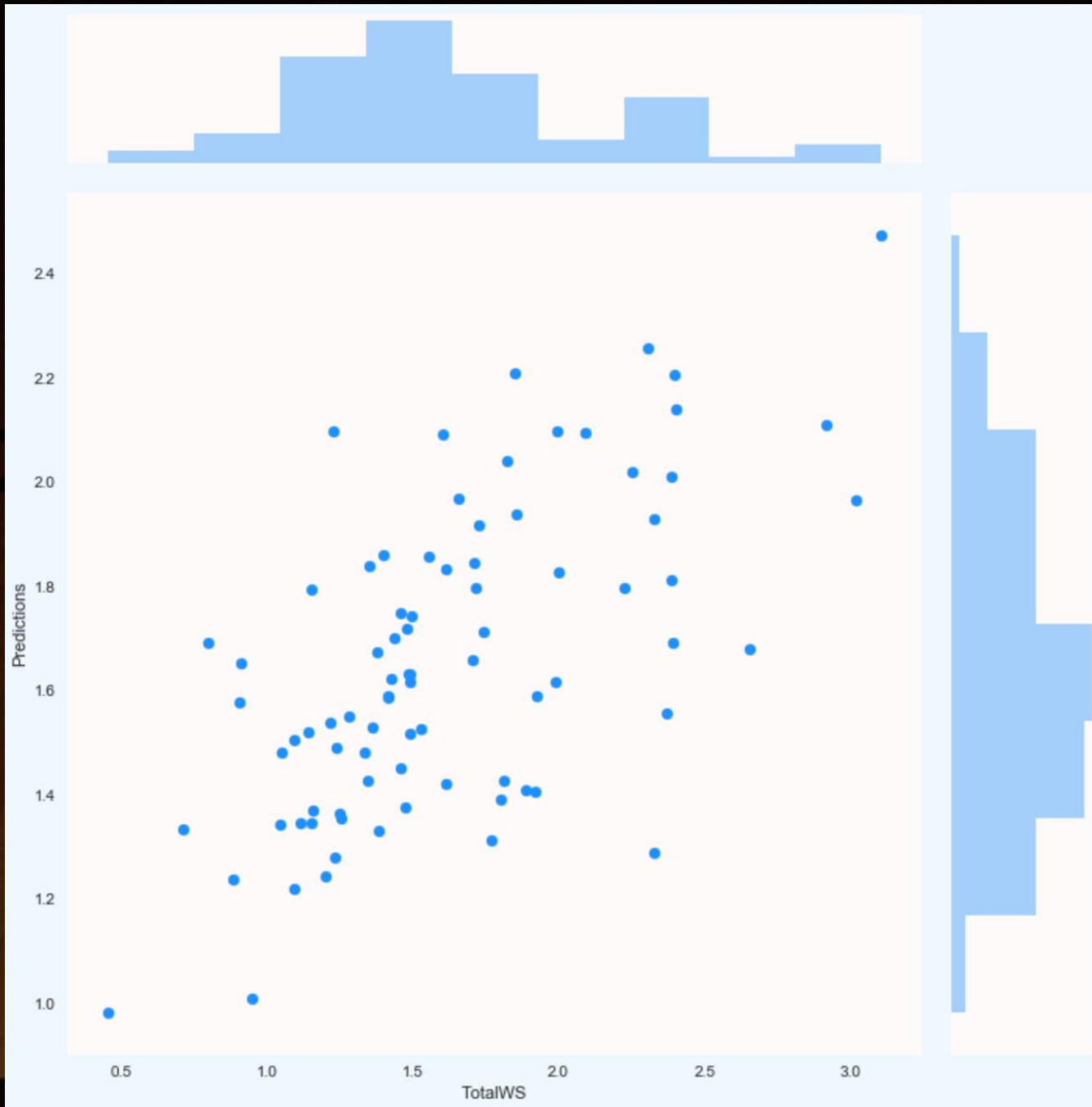
Overall Skill

RidgeCV with alpha of 0.1

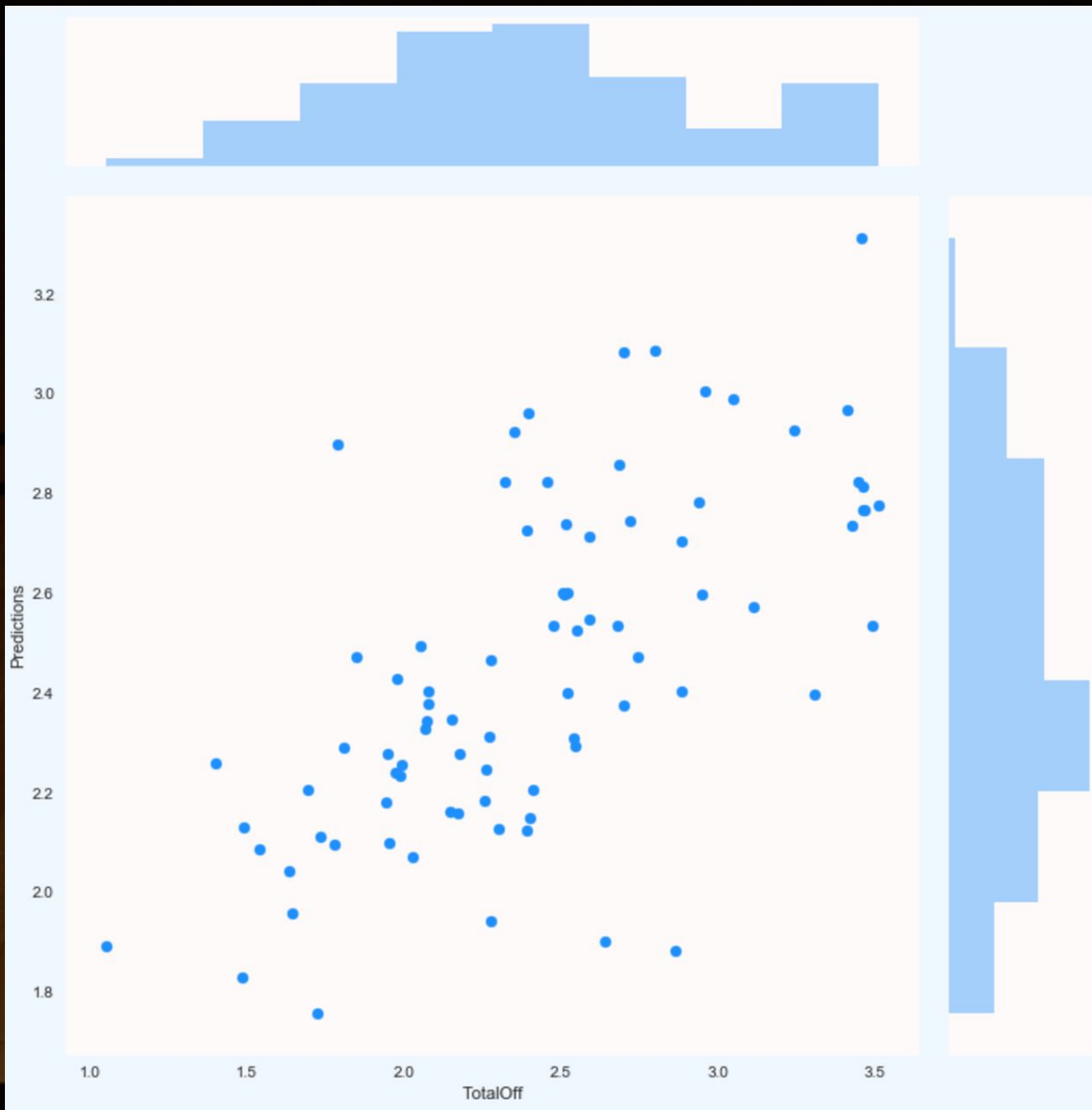|	Coefficients
FG%	1.277775
APGCR	0.822842
SPGSR	0.664581
FT%C	0.100024
PPG	0.088724
RPG	0.071291
TopSchool	0.069495
PF	0.039478
FTMSR	0.007872
Weight at Draft	0.000618
Height Inches C	0.000001
Age at Draft SR	-0.002522
BPG	-0.055338
3PMSR	-0.119240
FGM	-0.266576
TOVSR	-0.661566

Normalized Metric:
- Player Efficiency Rating (PER)
- Win Shares
- Box +/-
- Value Over Replacement Player (VORP)

Evaluation

R²: **38.5%**
RMSE: **0.41**

Offense

RidgeCV with alpha of 0.01

	Coefficients
FG%	3.725624e+00
ORBSR	3.351938e-01
TopSchool	1.862774e-01
SPGSR	1.859636e-01
FTMSR	1.685090e-01
PPG	9.995334e-02
3PMSR	5.408547e-02
BPG	1.322519e-02
Weight at Draft	1.286415e-03
Games Played 4	-7.450581e-08
Height Inches C	-5.559996e-07
Age at Draft SR	-2.225938e-03
PF	-1.765188e-01
3P%	-2.184724e-01
TOVSR	-2.273669e-01
FGM	-2.645271e-01
FT%C	-5.408843e-01

Normalized Metric:

- Effective Field Goal % (eFG%)
- True Shooting %
- Free Throw Rate
- Usage Rate
- Offensive Rebound %
- Offensive Win Shares

Evaluation

R²: 42.4%
RMSE: 0.43
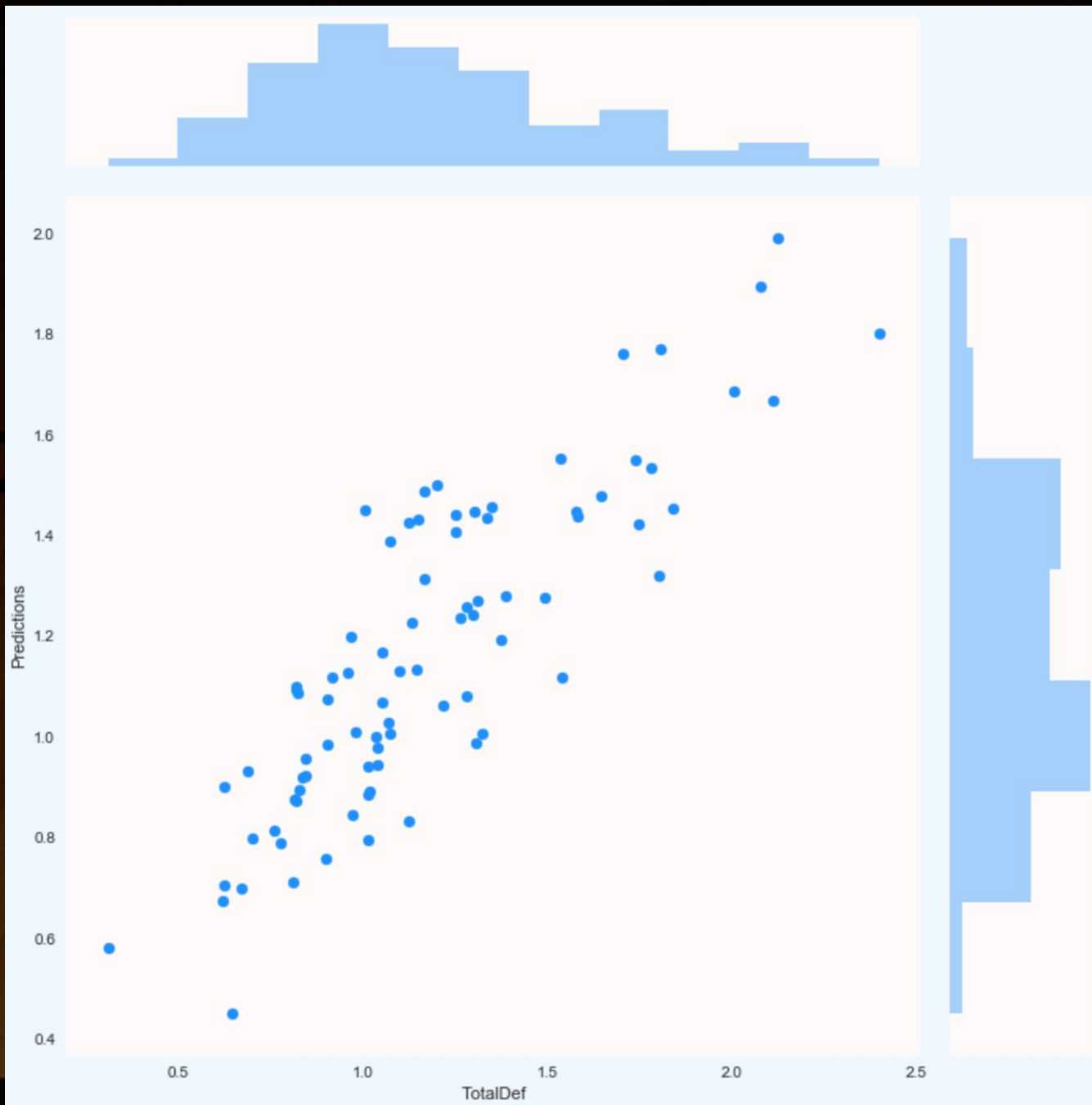
Defense

LassoCV with alpha of 0.001

	Coefficients
SPGSR	0.492733
DRBL	0.379234
APGCR	0.221876
FG%	0.178673
TopSchool	0.125770
BPG	0.036124
PF	0.026555
Weight at Draft	0.001159
Height Inches C	0.000001
FT%C	-0.000000
TOVSR	-0.000000
MPG3	-0.000006
Age at Draft SR	-0.001226
FTMSR	-0.122043
3PMSR	-0.235993

Normalized Metric:
- Defensive Rebound %
- Defensive Win Shares
- Defensive Box +/-

Evaluation

R²: 72.9%
RMSE: 0.21

Results

Test Class 1

Players who have not completed five seasons yet

Overall Skill

- Ranked by Ridge predictions

- Favors Field Goal %, Assists, and Steals

- Dislikes Free Throw %

- Notable Picks: Kyle Anderson (1), Marcus Smart (2), Joel Embiid (5)

	Model Pick	Year_x	Pick	Name	Team_x	PredictionsLR	PredictionsRG	PredictionsLS
0	1	2014	30	Kyle Anderson	SAS	3.418196	3.423443	3.385948
1	2	2014	6	Marcus Smart	BOS	3.353363	3.370434	3.353086
2	3	2014	18	Tyler Ennis	PHO	3.263149	3.271824	3.226756
3	4	2014	22	Jordan Adams	MEM	3.238085	3.229415	3.165659
4	5	2014	3	Joel Embiid	PHI	3.102121	3.030870	2.999040
5	6	2014	24	Shabazz Napier	CHH	2.982978	2.984984	2.969164
6	7	2014	7	Julius Randle	LAL	2.903000	2.927789	2.921210
7	8	2014	4	Aaron Gordon	ORL	2.897083	2.881866	2.876324
8	9	2014	59	Xavier Thames	TOR	2.721461	2.756754	2.733266
9	10	2014	39	Jerami Grant	PHI	2.698197	2.711564	2.694733
10	11	2014	19	Gary Harris	CHI	2.720959	2.710359	2.695052
11	12	2014	1	Andrew Wiggins	CLE	2.672674	2.689295	2.681367
12	13	2014	42	Nick Johnson	HOU	2.671553	2.682792	2.676700
13	14	2014	2	Jabari Parker	MIL	2.663643	2.680098	2.683570
14	15	2014	9	Noah Vonleh	CHH	2.695487	2.671400	2.677585
15	16	2014	48	Lamar Patterson	MIL	2.618828	2.648603	2.648076
16	17	2014	55	Semaj Christon	MIA	2.623185	2.646971	2.648882
17	18	2014	49	Cameron Bairstow	CHI	2.618868	2.636540	2.615931
18	19	2014	33	Joe Harris	CLE	2.616150	2.603201	2.601551
19	20	2014	14	T.J. Warren	PHO	2.634672	2.591980	2.583239
20	21	2014	44	Markel Brown	MIN	2.581216	2.587712	2.582719
21	22	2014	47	Russ Smith	PHI	2.540279	2.581620	2.579812
22	23	2014	13	Zach Lavine	MIN	2.565880	2.571391	2.543784
23	24	2014	17	James Young	BOS	2.519335	2.546913	2.550601
24	25	2014	15	Adreian Payne	ATL	2.507350	2.480219	2.475389
25	26	2014	26	P.J. Hairston	MIA	2.392007	2.438306	2.423502
26	27	2014	50	Alec Brown	PHO	2.376446	2.396661	2.389955
27	28	2014	32	K.J. Mcdaniels	PHI	2.366148	2.378455	2.379078
28	29	2014	11	Doug Mcdermott	DEN	2.360199	2.306040	2.315950
29	30	2014	37	Deandre Daniels	TOR	2.141392	2.185089	2.176867

Test Class 1

Players who have not completed five seasons yet

Offensive Skill

- Ranked by Ridge predictions

- Favors Field Goal %, Offensive Rebounds

- Dislikes Turnovers

- Notable Picks: Joel Embiid (1), Julius Randle (2), T.J. Warren (4)

	Model Pick	Year_y	Pick	Name	Team_x	PredictionsLR	PredictionsRG	PredictionsLS
0	1	2014.0	3	Joel Embiid	PHI	3.839349	3.812728	3.772134
1	2	2014.0	7	Julius Randle	LAL	3.783038	3.808807	3.797392
2	3	2014.0	22	Jordan Adams	MEM	3.609478	3.617696	3.612708
3	4	2014.0	14	T.J. Warren	PHO	3.609497	3.601414	3.558838
4	5	2014.0	49	Cameron Bairstow	CHI	3.507763	3.520280	3.539000
5	6	2014.0	9	Noah Vonleh	CHH	3.536972	3.518080	3.509918
6	7	2014.0	2	Jabari Parker	MIL	3.491009	3.499082	3.508858
7	8	2014.0	11	Doug Mcdermott	DEN	3.486192	3.471715	3.448604
8	9	2014.0	6	Marcus Smart	BOS	3.439455	3.441182	3.441099
9	10	2014.0	39	Jerami Grant	PHI	3.420280	3.436072	3.418010
10	11	2014.0	30	Kyle Anderson	SAS	3.409072	3.425937	3.455056
11	12	2014.0	15	Adreian Payne	ATL	3.368271	3.364188	3.348424
12	13	2014.0	4	Aaron Gordon	ORL	3.318918	3.339064	3.319676
13	14	2014.0	1	Andrew Wiggins	CLE	3.330036	3.338667	3.332850
14	15	2014.0	32	K.J. Mcdaniels	PHI	3.341092	3.337556	3.329310
15	16	2014.0	19	Gary Harris	CHI	3.218520	3.218879	3.213667
16	17	2014.0	17	James Young	BOS	3.181479	3.214155	3.210416
17	18	2014.0	50	Alec Brown	PHO	3.160067	3.170460	3.215516
18	19	2014.0	18	Tyler Ennis	PHO	3.099621	3.112759	3.141821
19	20	2014.0	44	Markel Brown	MIN	3.108146	3.109483	3.114961
20	21	2014.0	33	Joe Harris	CLE	3.096984	3.096899	3.098021
21	22	2014.0	24	Shabazz Napier	CHH	3.063094	3.067471	3.074162
22	23	2014.0	13	Zach Lavine	MIN	3.042642	3.063887	3.077694
23	24	2014.0	48	Lamar Patterson	MIL	3.042334	3.046467	3.083344
24	25	2014.0	55	Semaj Christon	MIA	3.031608	3.034655	3.072693
25	26	2014.0	42	Nick Johnson	HOU	3.009011	3.021431	3.039140
26	27	2014.0	26	P.J. Hairston	MIA	2.992480	3.014865	3.062396
27	28	2014.0	59	Xavier Thames	TOR	2.961504	2.971449	3.023363
28	29	2014.0	47	Russ Smith	PHI	2.960889	2.967563	2.985715
29	30	2014.0	37	Deandre Daniels	TOR	2.854423	2.868391	2.928017

Test Class 1

Players who have not completed five seasons yet

Defensive Skill

- Ranked by Lasso predictions

- Favors Steals, Rebounds

- Dislikes 3 Pointers

- Notable Picks: Joel Embiid (1), Kyle Anderson (2), Marcus Smart (6)

	Model Pick	Year_x	Pick	Name	Team_x	PredictionsLR	PredictionsRG	PredictionsLS	PredictionsGBR
0	1	2014	3	Joel Embiid	PHI	2.376027	2.354394	2.303524	1.764845
1	2	2014	30	Kyle Anderson	SAS	2.310856	2.321525	2.302164	1.640984
2	3	2014	7	Julius Randle	LAL	1.973079	1.989446	1.999910	1.465746
3	4	2014	4	Aaron Gordon	ORL	2.013344	2.010722	1.938391	1.637066
4	5	2014	9	Noah Vonleh	CHH	1.929634	1.928312	1.930446	1.356614
5	6	2014	6	Marcus Smart	BOS	1.856253	1.874224	1.879600	1.329858
6	7	2014	50	Alec Brown	PHO	1.767461	1.791697	1.801619	1.259719
7	8	2014	22	Jordan Adams	MEM	1.828386	1.828748	1.801350	1.005499
8	9	2014	2	Jabari Parker	MIL	1.768634	1.778761	1.788273	1.344793
9	10	2014	49	Cameron Bairstow	CHI	1.774164	1.789501	1.787392	1.291077
10	11	2014	15	Adreian Payne	ATL	1.794815	1.790693	1.774901	1.370220
11	12	2014	39	Jerami Grant	PHI	1.709802	1.724180	1.699912	1.214322
12	13	2014	14	T.J. Warren	PHO	1.784967	1.753398	1.689569	1.338170
13	14	2014	18	Tyler Ennis	PHO	1.634628	1.646633	1.606383	1.046782
14	15	2014	37	Deandre Daniels	TOR	1.543415	1.568554	1.593638	1.100531
15	16	2014	32	K.J. Mcdaniels	PHI	1.570393	1.584244	1.590047	1.043307
16	17	2014	48	Lamar Patterson	MIL	1.539860	1.564328	1.573125	1.202714
17	18	2014	13	Zach Lavine	MIN	1.540328	1.545112	1.511673	1.049838
18	19	2014	1	Andrew Wiggins	CLE	1.486239	1.503026	1.501612	1.024696
19	20	2014	24	Shabazz Napier	CHH	1.498687	1.506209	1.496633	1.022684
20	21	2014	42	Nick Johnson	HOU	1.490740	1.503535	1.493768	0.974222
21	22	2014	55	Semaj Christon	MIA	1.493177	1.499828	1.488304	1.013588
22	23	2014	44	Markel Brown	MIN	1.486258	1.496588	1.481060	1.097035
23	24	2014	17	James Young	BOS	1.440339	1.459180	1.453783	0.908757
24	25	2014	33	Joe Harris	CLE	1.456121	1.457815	1.430366	1.031492
25	26	2014	19	Gary Harris	CHI	1.381651	1.380013	1.362046	0.781719
26	27	2014	47	Russ Smith	PHI	1.316104	1.336203	1.356820	1.181879
27	28	2014	11	Doug Mcdermott	DEN	1.424018	1.392152	1.351678	1.107913
28	29	2014	59	Xavier Thames	TOR	1.307010	1.333902	1.349193	0.897125
29	30	2014	26	P.J. Hairston	MIA	1.180460	1.214219	1.254089	0.851989

Test Class 2

Players who declared as early listing for the 2017 NBA Draft

Overall Skill

- Ranked by Ridge predictions

- Favors Field Goal %, Assists, and Steals

- Dislikes Free Throw %

	Model Pick	Name	Team	PredictionsLR	PredictionsRG	PredictionsLS
0	1	Lonzo Ball	UCLA	3.522445	3.397936	3.367442
1	2	Nigel Williams-Goss	GU	3.236383	3.199645	3.153059
2	3	Tyler Lydon	SU	3.023141	2.988966	2.954506
3	4	Dennis Smith	NCST	3.002844	2.978863	2.984050
4	5	Jonathan Isaac	FSU	3.011498	2.969870	2.939302
5	6	De'Aaron Fox	UK	2.928345	2.935485	2.894510
6	7	Alize Johnson	MSU	2.960759	2.931956	2.942168
7	8	Jawun Evans	OSU	2.912080	2.904007	2.872870
8	9	Ivan Rabb	Cal	2.930605	2.901922	2.883669
9	10	Josh Jackson	KU	2.915572	2.878590	2.881542
10	11	Jayson Tatum	DUKE	2.826655	2.834493	2.808966
11	12	Lauri Markkanen	U of A	2.853445	2.814047	2.787789
12	13	Melo Trimble	UMD	2.807886	2.806628	2.799442
13	14	Luke Kennard	DUKE	2.811750	2.789817	2.747077
14	15	Markelle Fultz	UW	2.800023	2.771351	2.778537
15	16	Isaiah Briscoe	UK	2.742836	2.767429	2.754398
16	17	Caleb Swanigan	PU	2.764380	2.743554	2.757687
17	18	Eric Mika	BYU	2.719254	2.738978	2.715722
18	19	Andrew Jones	UT	2.730748	2.734914	2.727355
19	20	Vince Edwards	PU	2.746294	2.712348	2.685351
20	21	Thomas Bryant	IU	2.788775	2.703679	2.689664
21	22	Tony Bradley	UNC	2.737763	2.698408	2.668401
22	23	Aaron Holiday	UCLA	2.644252	2.643830	2.624489
23	24	Justin Patton	CU	2.763796	2.641447	2.628204
24	25	Chandler Hutchison	BSU	2.591501	2.614697	2.604113
25	26	Rawle Alkins	U of A	2.629015	2.614630	2.611058
26	27	Edmond Sumner	XAV	2.533838	2.607258	2.596028
27	28	Zach Collins	GU	2.719354	2.603178	2.584694
28	29	Dillon Brooks	UO	2.605569	2.593719	2.572262
29	30	Justin Jackson	UMD	2.556794	2.564600	2.574161
30	31	L.J. Peak	GU	2.557903	2.561064	2.549566
31	32	Justin Jackson	UNC	2.567257	2.547657	2.541250

Test Class 2

Players who declared as early listing for the 2017 NBA Draft

Offensive Skill

- Ranked by Ridge predictions

- Favors Field Goal %, Offensive Rebounds

- Dislikes Turnovers

	Model Pick	Name	Team	PredictionsLR	PredictionsRG	PredictionsLS
0	1	Tacko Fall	UCF	4.094820	4.053926	3.936276
1	2	Tony Bradley	UNC	3.767375	3.774038	3.709878
2	3	John Collins	WF	3.726176	3.717319	3.688870
3	4	Lonzo Ball	UCLA	3.712867	3.688546	3.590400
4	5	Zach Collins	GU	3.690312	3.674693	3.595213
5	6	Ike Anigbogu	UCLA	3.669028	3.660945	3.634125
6	7	Jarrett Allen	UT	3.658016	3.650747	3.619150
7	8	Eric Mika	BYU	3.634698	3.644818	3.643473
8	9	Thomas Bryant	IU	3.617770	3.593965	3.542936
9	10	Justin Patton	CU	3.608082	3.585030	3.525115
10	11	Harry Giles	DUKE	3.583898	3.566108	3.518230
11	12	Lauri Markkanen	U of A	3.487395	3.494783	3.463933
12	13	Ivan Rabb	Cal	3.499870	3.494644	3.491740
13	14	Jonathan Isaac	FSU	3.461938	3.453674	3.423392
14	15	Alize Johnson	MSU	3.466398	3.451814	3.405759
15	16	Drew Eubanks	OSU	3.460975	3.450437	3.450763
16	17	Markelle Fultz	UW	3.402481	3.381067	3.367000
17	18	Dennis Smith	NCST	3.369375	3.354302	3.325703
18	19	Kyle Kuzma	U of U	3.347372	3.336992	3.312944
19	20	Josh Jackson	KU	3.337796	3.331672	3.292311
20	21	De'Aaron Fox	UK	3.285145	3.299914	3.316495
21	22	Isaac Humphries	UK	3.243443	3.275859	3.351652
22	23	Devin Robinson	UF	3.250455	3.266167	3.274294
23	24	Tyler Dorsey	UO	3.260918	3.261739	3.219612
24	25	Malik Monk	UK	3.241593	3.258453	3.271921
25	26	Isaiah Briscoe	UK	3.237598	3.249922	3.276033
26	27	Caleb Swanigan	PU	3.233963	3.225725	3.239856
27	28	Tyler Lydon	SU	3.216322	3.221701	3.227785
28	29	Justin Jackson	UMD	3.213978	3.214996	3.219379
29	30	Nigel Williams-Goss	GU	3.197595	3.205393	3.215334
30	31	Antonio Blakeney	LSU	3.191733	3.186771	3.178502
31	32	Luke Kennard	DUKE	3.168374	3.178150	3.192581

Test Class 2

Players who declared as early listing for the 2017
NBA Draft

Defensive Skill

- Ranked by Lasso predictions

- Favors Steals, Rebounds

- Dislikes 3 Pointers

	Model Pick	Name	Team	PredictionsLR	PredictionsRG	PredictionsLS	PredictionsGBR
0	1	Tacko Fall	UCF	2.436848	2.377179	2.295436	1.784391
1	2	Justin Patton	CU	2.251647	2.188036	2.070089	1.704461
2	3	Drew Eubanks	OSU	2.025960	2.016663	1.999796	1.556603
3	4	Jonathan Isaac	FSU	2.013604	2.007477	1.967949	1.221629
4	5	Josh Jackson	KU	2.045177	2.027523	1.967795	1.704601
5	6	Jarrett Allen	UT	1.981672	1.974633	1.963342	1.642300
6	7	Ivan Rabb	Cal	2.009082	2.003923	1.952168	1.517764
7	8	Lonzo Ball	UCLA	2.137149	2.080677	1.950275	1.500709
8	9	Caleb Swanigan	PU	1.887175	1.885508	1.880217	1.485655
9	10	Zach Collins	GU	1.985396	1.942578	1.878404	1.520818
10	11	Tyler Lydon	SU	1.888438	1.891080	1.854413	1.281676
11	12	Eric Mika	BYU	1.849977	1.863760	1.848294	1.416098
12	13	Thomas Bryant	IU	1.928366	1.897890	1.842012	1.661645
13	14	John Collins	WF	1.863624	1.857811	1.837557	1.525581
14	15	Isaiah Briscoe	UK	1.854012	1.865413	1.836386	1.402552
15	16	Cameron Oliver	UNR	1.853281	1.855801	1.833090	1.453914
16	17	Jayson Tatum	DUKE	1.810436	1.823192	1.825364	1.275339
17	18	Omer Yurtseven	NCST	1.760827	1.795317	1.804072	1.274920
18	19	Ike Anigbogu	UCLA	1.828081	1.824914	1.793149	1.596065
19	20	De'Aaron Fox	UK	1.829456	1.830089	1.792506	1.119320
20	21	Harry Giles	DUKE	1.860721	1.844366	1.785431	1.560474
21	22	Alize Johnson	MSU	1.824868	1.818613	1.775502	1.212149
22	23	Nigel Williams-Goss	GU	1.823842	1.808400	1.766652	1.267696
23	24	Chandler Hutchison	BSU	1.711117	1.725540	1.710941	1.105703
24	25	Tony Bradley	UNC	1.769741	1.757666	1.703780	1.426114
25	26	Devin Robinson	UF	1.677288	1.692925	1.679844	1.045584
26	27	Justin Jackson	UMD	1.655124	1.667717	1.670666	1.152833
27	28	Kyle Kuzma	U of U	1.685036	1.679671	1.649418	1.483910
28	29	Johnathan Motley	BU	1.559365	1.611443	1.649137	1.371557
29	30	Andrew Jones	UT	1.627989	1.636493	1.637512	0.923375
30	31	Isaac Humphries	UK	1.551227	1.591821	1.619757	1.208236
31	32	Dennis Smith	NCST	1.642674	1.632708	1.600677	1.095945

Further Analysis

- Work with more data, especially demographic data

- Improve data cleaning process to increase sample size

- Utilize ensemble methods to incorporate predictions from different models

- Update and retrain model after every season and create pipeline that will generate predictions for incoming college players of that year

Thanks!

Q&A

Data Sources
- Basketball Reference
- Sports Reference
- DraftExpress
- RealGM
- NBA