



# The Perfect Draft



By James Hyman, Cameron Mitchell,  
and Nick Riccardi



# Background Info/Problem Recap

---

- NBA Draft held once per year after completion of each season
- Best draft eligible players selected by teams
  - Draft order determined by lottery: worst teams from previous season have best odds at getting first pick
- Problem: Hundreds of draft eligible players each year
- Can be hard to analyze all players
  - What stats most important, take into account team need

# Proposed Solution

---

- Cluster players into various groups based on statistical performance
  - Can be used to evaluate based off of team need
- Create predictive model based on player performance
  - Used to evaluate best overall player
- Combine two together with model/random forest/decision tree
- Only focus on analyzing players drafted in first round
  - More clarity, less randomness with players drafted in first round

# Dataset Overview: Players

- 12000 NCAA players dating back to 2009-10 Season
- Features: player name, team, season, GP, 20 player per game performance measures

	Player	Team	GP	MPG	FGM	FGA	FG%	3PM	3PA	3P%	...	TOV	PF	ORB	DRB	RPG	APG	SPG	BPG	PPG	season
0	Aubrey Coleman	U of H	35	36.9	8.7	20.5	0.425	1.5	4.6	0.317	...	2.3	2.2	2.4	5.0	7.4	2.6	2.7	0.2	25.6	2010
1	Adnan Hodzic	LIP	30	32.3	9.0	14.8	0.604	0.0	0.0	0.000	...	2.7	2.2	3.4	5.6	9.1	0.8	0.5	0.4	22.7	2010
2	Keith Haynes	UTA	30	33.3	7.2	14.8	0.483	2.1	5.1	0.409	...	3.6	2.1	0.8	3.5	4.3	3.7	1.5	0.3	22.6	2010
3	Adrian Oliver	SJSU	31	36.1	7.4	17.0	0.437	1.6	4.0	0.408	...	3.2	3.0	1.6	3.7	5.3	2.9	1.0	0.5	22.5	2010
4	Devan Downey	USC	31	34.0	7.6	19.1	0.400	2.2	6.5	0.342	...	3.5	2.3	0.5	2.7	3.3	3.5	2.7	0.0	22.5	2010

# Dataset Overview: Teams

- Each NBA team from 2010-2020
- Features: WIN%, PTS, FG%, 3PT%, FT%, RPG, APG, TOV, STL, BLK

	TEAM	Season	GP	W	L	WIN%	MIN	PTS	FGM	FGA	...	DREB	REB	AST	TOV	STL	BLK	BLKA	PF	PFD	+/-
0	Utah Jazz	2020.0	72.0	52.0	20.0	0.722	48.2	116.4	41.3	88.1	...	37.6	48.3	23.7	14.2	6.6	5.2	3.9	18.5	19.0	9.3
1	Phoenix Suns	2020.0	72.0	51.0	21.0	0.708	48.6	115.3	43.3	88.3	...	34.2	42.9	26.9	12.5	7.2	4.3	3.6	19.1	18.0	5.8
2	Philadelphia 76ers	2020.0	72.0	49.0	23.0	0.681	48.4	113.6	41.4	86.9	...	35.0	45.1	23.7	14.4	9.1	6.2	4.7	20.2	21.0	5.6
3	Brooklyn Nets	2020.0	72.0	48.0	24.0	0.667	48.3	118.6	43.1	87.3	...	35.5	44.4	26.8	13.5	6.7	5.3	4.6	19.0	18.9	4.5
4	Denver Nuggets	2020.0	72.0	47.0	25.0	0.653	48.6	115.1	43.3	89.2	...	33.9	44.4	26.8	13.5	8.1	4.5	4.5	19.1	19.2	4.9

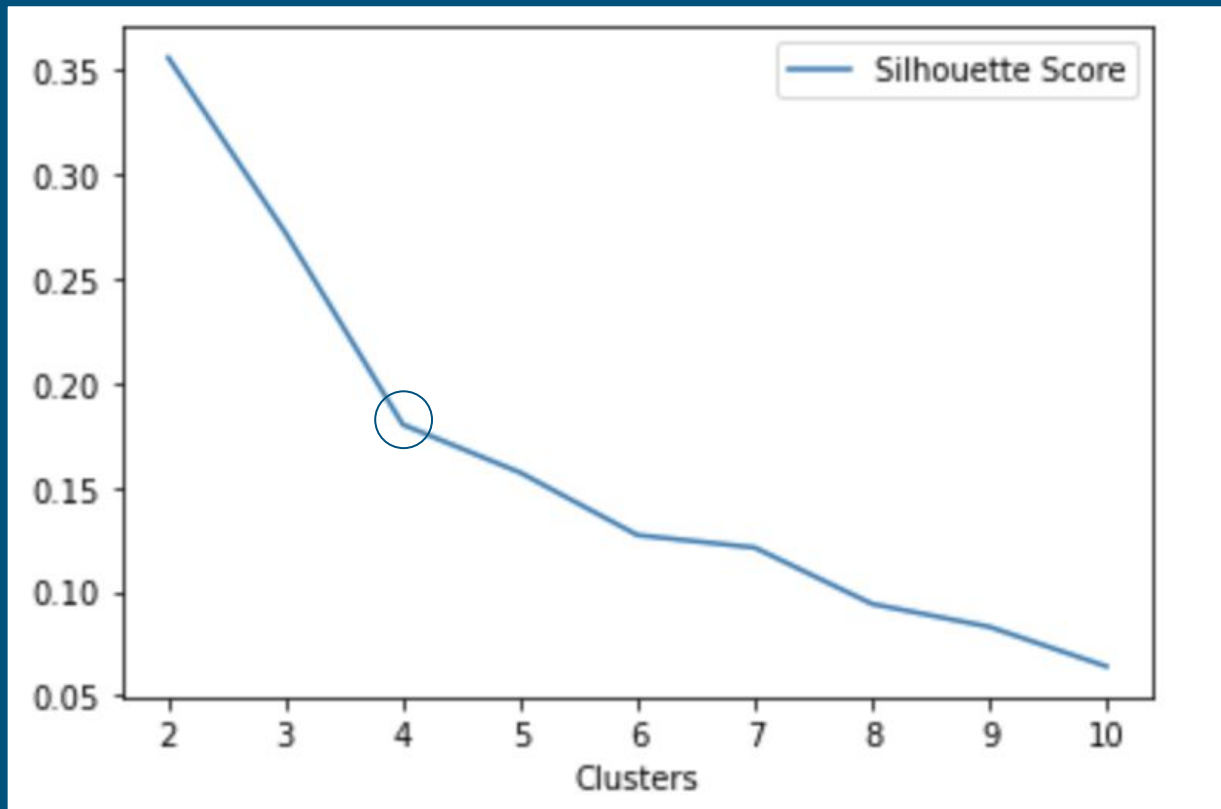
# Approach #1: Clustering

---

- Used k-means clustering, pipeline models in spark to create clusters
  - Clustered based on all player metrics mentioned earlier
- Clustered entire dataset, then filtered out first rounders for analysis
- Based on Silhouette score analysis, 4 clusters gave best results
- Although Silhouette scores were kind of low, clusters still appeared to be fairly defined, easy to identify groups

# Approach #1: Clustering

---



# Approach #1: Clustering

---

- Clustered Resulted in four groups of players who fall into various groups

Cluster	Skillset
0	Big Men
1	Scorers
2	3-Point Specialists
3	Defensive Specialists



# Approach #1: Clustering

cluster	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%	ORB	DRB	RPG	APG	SPG	BPG	PPG
0	5.53111111	10.1588889	0.54675556	0.44111111	1.25666667	0.26444444	3.37	4.93555556	0.67986667	2.82444444	5.9	8.72222222	1.53888889	0.97111111	1.58222222	14.8733333
1	6.35779817	13.7009174	0.46538532	1.95779817	5.17522936	0.37438532	4.20825688	5.29724771	0.79544037	1.29174312	4.33486239	5.62844037	3.21651376	1.36238532	0.54770642	18.8844037
2	4.38571429	9.61904762	0.45728571	1.72857143	4.45714286	0.39157143	1.95714286	2.55714286	0.76928571	0.88571429	3.20952381	4.10952381	2.11904762	0.95238095	0.46190476	12.4619048
3	4.66285714	10.2771429	0.45465714	1.25428571	3.42285714	0.35745714	3.01714286	4.1	0.74217143	1.00285714	3.69142857	4.68857143	4.00857143	1.47714286	0.49428571	13.5914286

# Approach #1: Clustering

---

- Cluster 0: Big Guys
  - Highest Rebounds/Game, Blocks/Game, Field Goal%
- Example: Anthony Davis



- Cluster 1: Scorers
  - Highest points/game, field goal makes/attempts, 2nd FG%
- Example: Cade Cunningham



# Approach #1: Clustering

---

- Cluster 2: Shooters
  - Highest 3P%, fewest free throws made/attempted
- Example: Landry Shamet



- Cluster 3: Defensive Guys
  - Highest Steals/game
- Example: Eric Bledsoe



# Approach #1: Clustering

---

- 

Cluster	Average Draft Position
0	14
1	15
2	18
3	17

# Approach #2: Linear Regression

---

- Created linear regression model to predict draft pick
- Used as a rating system for players
- Features include player stats and mock draft data

# Approach #2: Linear Regression

- Features: PPG, FGM, 3PM, RPG, TOV, BPG, APG, FGA, MPG, FTM, TOV, Mock
- R-squared (validation): 0.684
- RMSE (testing): 5.22

OVERALL PICK	Drafted	Mock	features	prediction
1.0	1.0	1.0	[20.1,6.5,2.3,6.2...	3.641299775898841
3.0	1.0	3.0	[16.4,6.0,0.4,8.7...	4.756561558513077
5.0	1.0	4.0	[14.4,5.2,1.2,5.3...	5.554537805462189
4.0	1.0	5.0	[10.3,4.1,0.5,4.0...	6.512639348763637
14.0	1.0	8.0	[16.8,5.2,1.8,5.8...	8.643664221477822
8.0	1.0	9.0	[12.5,4.4,1.2,6.5...	11.411223814971148
15.0	1.0	13.0	[18.6,6.5,2.8,5.0...	12.348449393331093
9.0	1.0	14.0	[14.0,5.3,2.1,2.7...	13.124978181798967
13.0	1.0	15.0	[17.1,6.2,2.3,4.6...	13.253801970689448
20.0	1.0	16.0	[16.5,6.9,0.0,7.5...	15.034270903558486

# Team Needs

---

- Teams will draft players based on what they are missing
- Original goal was to use a Random Forest model to predict type of player needed most (Shooter, Big Man, Scorer, Defense) based on previous year's stats
- Switched to Multinomial Logistic Regression

# Team Needs: Multinomial Logistic Regression

- Focus on team stats that indicate weaknesses (3P%, Reb, Steals, etc.)
- Training Accuracy: 47.5
- Testing Accuracy: 46.1

predicted_cluster	0.0	1.0	2.0	3.0
cluster				
0.0	29	57	1	2
1.0	21	84	1	2
2.0	3	12	2	4
3.0	5	26	0	4

- Pick location a major factor, scorers will be valued by all teams



# Draft Comparison

Pick	▼	Actual	▼	Ours	▼	Pick	▼	Actual	▼	Ours	▼
1		Cade Cunningham		Cade Cunningham		16		Alperen Sengun		Ziaire Williams	
2		Jalen Green		Evan Mobley		17		Trey Murphy III		Bones Hyland	
3		Evan Mobley		Jalen Suggs		18		Tre Mann		Burile Belo	
4		Scottie Barnes		Scottie Barnes		19		Kai Jones		Nick Muszynski	
5		Jalen Suggs		Moses Moody		20		Jalen Johnson		Ayo Dosunmu	
6		Josh Giddey		Franz Wagner		21		Keon Johnson		Scott Blakney	
7		Jonathan Kuminga		Corey Kispert		22		Isaiah Jackson		Drew Timme	
8		Franz Wagner		Davion Mitchell		23		Usman Garuba		Jaden Springer	
9		Davion Mitchell		Chris Duarte		24		Josh Christopher		D.J. Burns, Jr.	
10		Ziaire Williams		Jalen Johnson		25		Quentin Grimes		Scotty Pippen, Jr.	
11		James Bouknight		Trey Murphy III		26		Bones Hyland		Braelen Bridges	
12		Joshua Primo		Keon Johnson		27		Cam Thomas		James Jean-Marie	
13		Chris Duarte		Miles McBride		28		Jaden Springer		Matt Lewis	
14		Moses Moody		Jared Butler		29		Day'Ron Sharpe		Oscar da Silva	
15		Corey Kispert		Cam Thomas		30		Santi Aldama		Karlis Silins	

# Factors to Improve Results

---

- Controllables

1. Including International Players

2. Gathering more advanced metrics

- Uncontrollables

1. Player Character

2. Injury History

---

Thank you  
Questions?