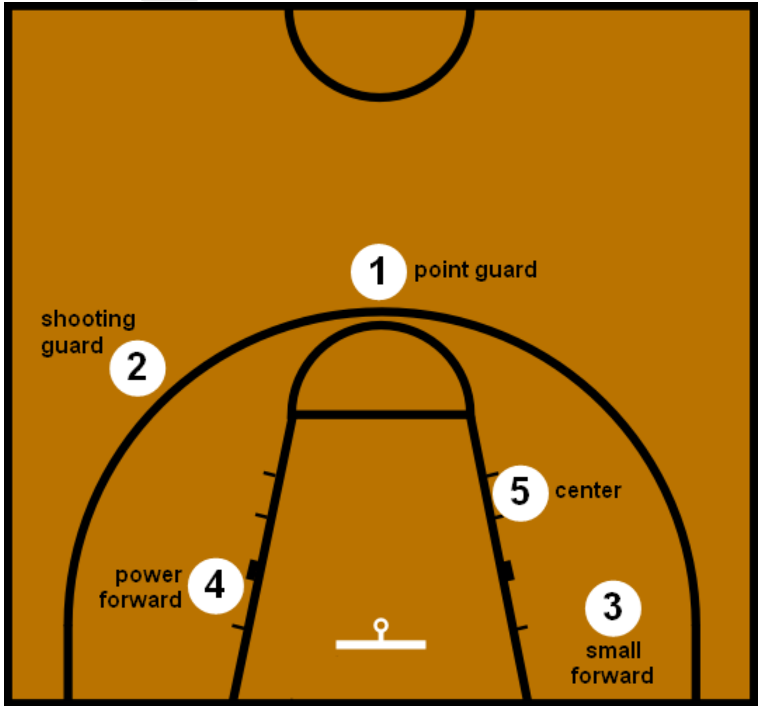
Classification of NBA players

Zibo Song & Tao wang

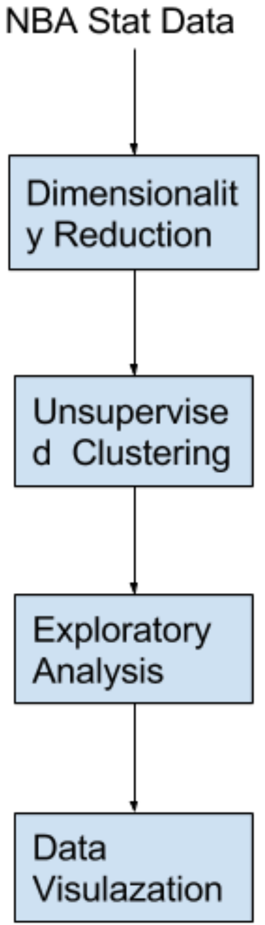
Introduction & Motivation



- The traditional five player positions incorrectly oversimplify the skill sets of NBA players. Simply pigeon-holing players into one of five positions does not accurately define a player's specific skill set.

- To uncover the positions that are intrinsic to today's NBA players and classify players with a position that best encapsulates their skill sets.

Proposed Solution



- Data Collection

We scrape our data from Basketball-Reference.com

Player Per 100 Poss

Share & more ▼ Glossary Hide Partial Rows

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	ORtg	DRtg
1	Alex Abrines	SG	23	OKC	68	6	1055	6.2	15.9	.393	4.4	11.5	.381	1.9	4.4	.426	2.0	2.3	.898	0.8	3.2	4.0	1.9	1.7	0.4	1.5	5.3	18.9	113	110

Advanced

Share & more ▼ Glossary Hide Partial Rows

Rk	Player	Pos	Age	Tm	G	MP	PER	TS%	3PAr	FTr	ORB%	DRB%	TRB%	AST%	STL%	BLK%	TOV%	USG%	OWS	DWS	WS	WS/48	OBPM	DBPM	BPM	VORP
1	Alex Abrines	SG	23	OKC	68	1055	10.1	.560	.724	.144	1.9	7.1	4.5	5.5	1.7	0.6	8.3	15.9	1.2	0.9	2.1	.096	-0.3	-2.2	-2.5	-0.1

To better define a player, we combine per-100 possesions and advance metrics

Prior to analysis, the data consisted of 486 players and 45 features (or dimensions) from 2016 season

- Dimensionality Reduction

In this project, each dimension is represented by a player’s feature statistics (i.e. PER, TS%, 3P%, etc.) and in order to obtain a statistically sound result, the amount of data should better be reduced by obtaining a set of principal components.

For data visualization reasons, we choose to use 2 principal components.

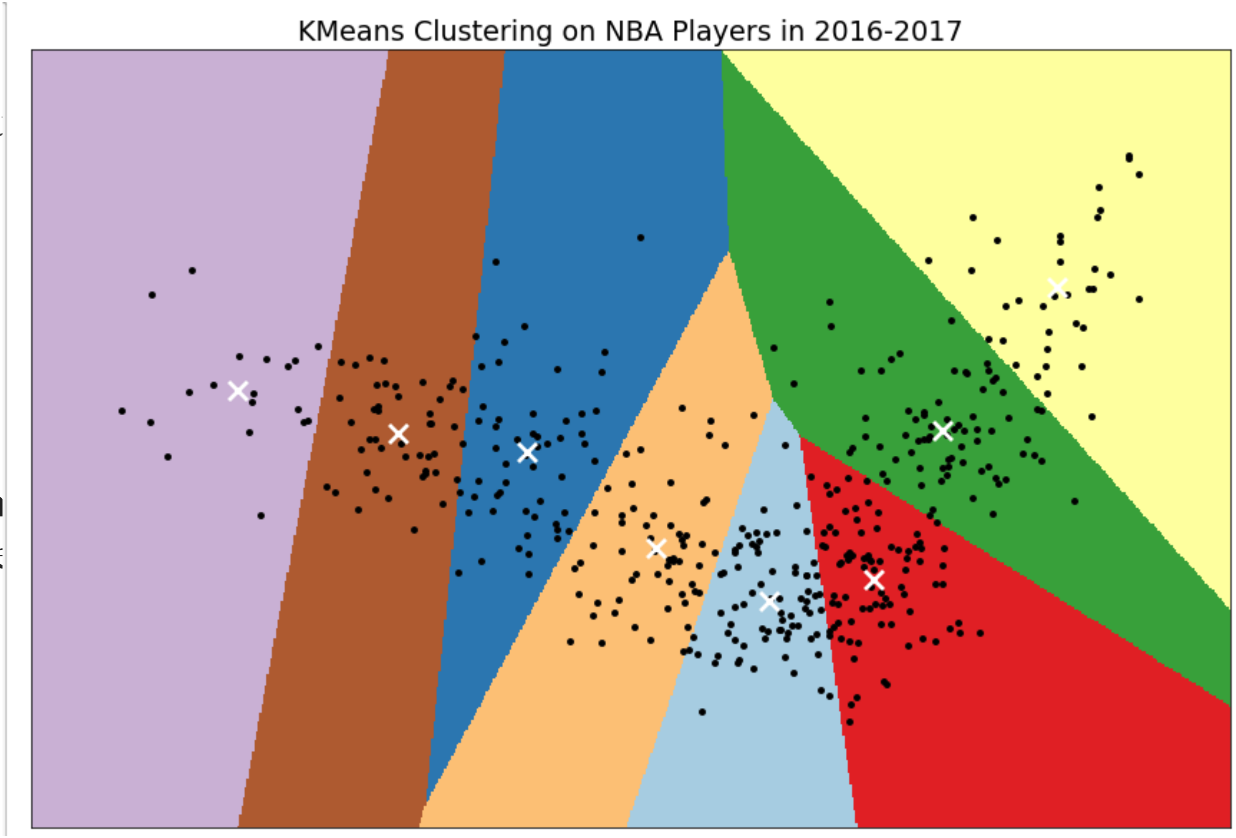
Then we compare two dimensionality reduction methods: PCA and LDA(Linear Discriminant Analysis)

PCA: Capture 49.43% of the data with two components

LDA: Capture 65.27% of the data with two components

- Cluster The Data with K-means

Here we use silhouette score to decide K.



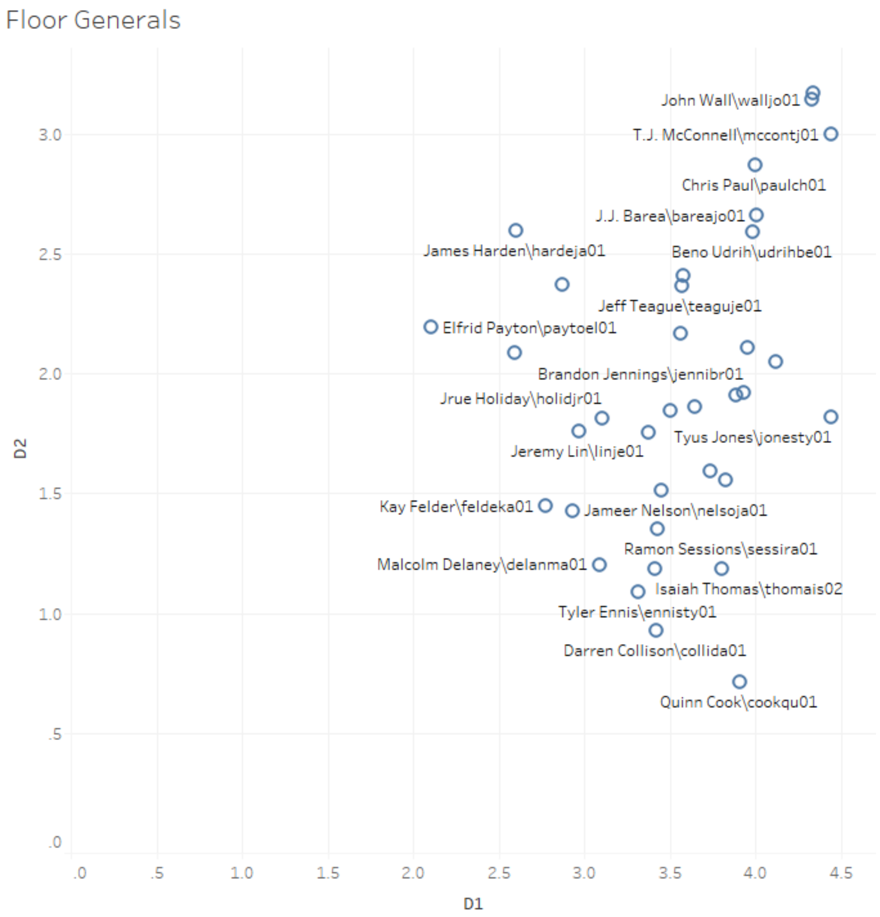
clu ther

Th ed all

the

Then, we use PCA to identify the most important features so that we can define each cluster in common words.

1. Versatile Center
2. Scoring Wings
3. Versatile Forwards
4. Offensive Centers
5. Floor Generals
6. Shooting Wings
7. Defensive Centers
8. 3-and-D Wings



Notable Player: James Harden, Chris Paul, John Wall

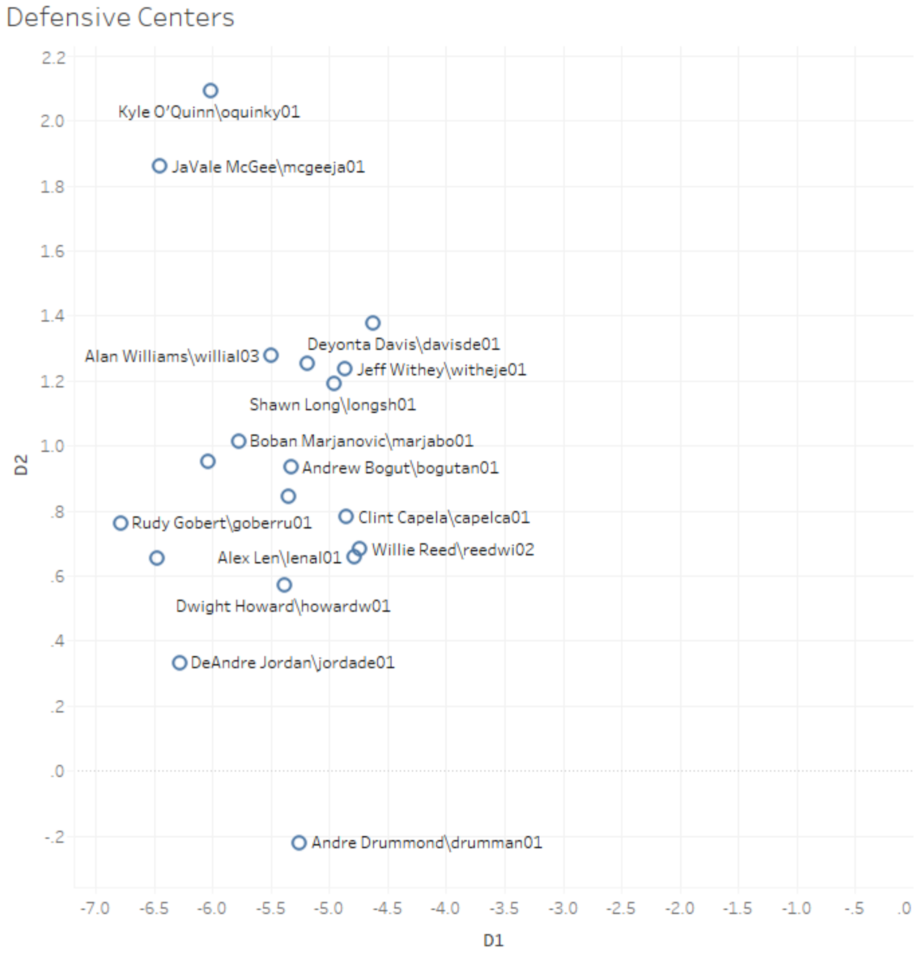
10 most important features for cluster
----> Floor General

	Feature	Importance	Cluster Average	League Average
0	BPM	0.223549	-0.991429	-1.155631
1	PER	0.221773	15.282857	13.533559
2	VORP	0.219615	0.954286	0.677252
3	WS	0.214771	3.545714	2.822297
4	OBPM	0.211280	0.348571	-0.995495
5	WS/48	0.208305	0.087771	0.085556
6	OWS	0.204775	2.188571	1.464865
7	PTS	0.191846	21.617143	19.915090
8	FT	0.189831	3.865714	3.227703
9	FTA	0.187174	4.760000	4.298198

BPM: Box Plus/Minus Value

PER: Player Efficiency Rating

VORP: Value Over Replacement Player



10 most important features for cluster - defensive centers

	Feature	Importance	Cluster Average	League Average
0	TOV%	0.172287	14.157895	12.789414
1	PF	0.162561	6.542105	4.474775
2	DRtg	0.122652	103.052632	108.979730
3	DBPM	0.098983	2.136842	-0.162162
4	AST	0.071157	2.036842	4.261261
5	AST%	0.045135	6.368421	13.215991
6	3P%	0.025636	0.084474	0.287741
7	BLK%	0.023664	4.678947	1.666892
8	TOV	0.018149	2.615789	2.666667
9	STL%	0.002427	1.373684	1.563063

TOV%: Turnover Percentage

PF: Personal Foul

DRtg: Defensive Rating

Conclusion

- The clusters that our algorithms have constructed identify which features are most important to a player and group them in such a way that is easily interpretable

- Trends in the NBA are constantly changing and this study was intended to provide just a snapshot of today's NBA players.