

The Next Super Star in NBA?!

Yonglin Jing

Mengxin Ji

Meng Li

Ruolai Wu

Why this topic?

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

Given the large basis of NBA audiences and increasing pursue for super stars, using past year data to analyze who is the next super star seems to be an interesting and meaningful question.

- Are there any statistical patterns in the players performance? Can we use the pattern to predict the potential stars?
- What features do the potential stars and super stars share?

Methodology

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

Using NBA statistic data, we would like to do some exploratory analysis and data visualization over “all stars” and “potential players” based on their first four seasons data

- 1 Data Scrapping
- 2 K-means Cluster
- 3 PCA and Data Visualization
- 4 Compare player A and player B

Data Source

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

1 No API, use lxml module for web scrapping

The screenshot shows the Basketball Reference website. At the top, there's a navigation bar with links for Sports Reference, Baseball, Football (college), Basketball (college), Hockey, Soccer, Blog, Questions or Comments?, Create Account, and Login. Below this is the Basketball Reference logo and a search bar. A secondary navigation bar includes links for Players, Teams, Seasons, Leaders, Scores (with a red notification bubble), Playoffs, Draft, Play Index, and Full Site Menu Below. The main content area is divided into three columns. The left column, titled 'Every Player', shows a grid of player headshots and a section for 'View any Active Player' with a dropdown menu and a 'Go!' button. The middle column, titled 'Every Team', displays the '2016-17 NBA Standings' with a table showing the top 15 teams in the Eastern Conference. The right column, titled 'What's Happening', lists various news items including the '2016-17 NBA MVP Award Tracker', 'Playoff Probabilities Report', 'Projected Draft Order', 'Trending Player Pages', 'Hoops Rumors', and 'Sports Reference Blog'.

2016-17 NBA Standings
[Summary](#) [Schedule](#)

East	W	L
CLE (1)	42	19
BOS (2)	40	22
WAS (3)	36	24
TOR (4)	37	26
ATL (5)	34	27
IND (6)	31	30
CHI (7)	31	31
DET (8)	30	32
MIA (9)	29	34
MIL (10)	28	33
CHO (11)	27	35
NYK (12)	25	37
ORL (13)	23	39
PHI (13)	23	39
BKK (15)	10	51

Figure: Basketball Reference, <http://www.basketball-reference.com/>

Data Description

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

We combined the names of 84 players and the statistics of those players in four years into one data frame. In those 84 players, 35 of them are rookie players and 49 of them are stars. The data frame has 84 observations and 187 variables. Each variable is the average of the statistics in one year. For example, FGA_1 is the average of Field Goal Attempt (FGA) in year one of 84 players.

Age_1	G_1	GS_1	MP_1	FG_1	FGA_1	FG%_1	3P_1	3PA_1	3P%_1	...	USG%_4	OWS_4	DWS_4	WS_4	WS/48_4	OBPM_4	DBPM_4	BPM_4
21	82	79	36.7	6.7	13.0	.518	0.0	0.1	.200	...	26.2	4.1	2.8	6.8	.184	1.6	2.2	3.8
24	82	75	30.7	4.3	8.1	.530	0.0	0.0	.000	...	19.1	4.2	4.0	8.2	.166	0.3	3.6	3.9
19	82	82	32.6	4.3	8.3	.520	0.0	0.0	.000	...	24.2	6.4	6.4	12.9	.200	-0.2	2.5	2.2
19	77	72	29.4	3.5	8.3	.419	0.8	2.5	.323	...	25.1	3.6	4.5	8.1	.141	1.9	0.5	2.3
22	74	31	20.7	2.4	5.0	.482	0.0	0.0	.000	...	16.8	3.1	3.6	6.7	.205	0.7	4.2	4.9

Figure: Statistics of Players

K-means Cluster

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

Cluster 4 is the largest proportion and cluster 1 is the smallest. The players in cluster 2 are mostly rookie players and all the players in cluster 3 are stars. In cluster 1, 4, and 3, there are a few players are rookie players. Such that rookie stars in cluster 1, 4, and 3 may be potential candidates for next stars.

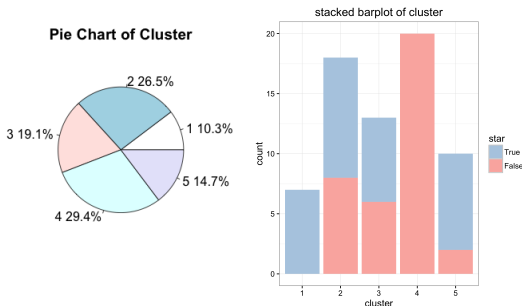


Figure: Pie plots of the Cluster, Stack Barplot of Cluster

Data Visualization

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally

Conclusion

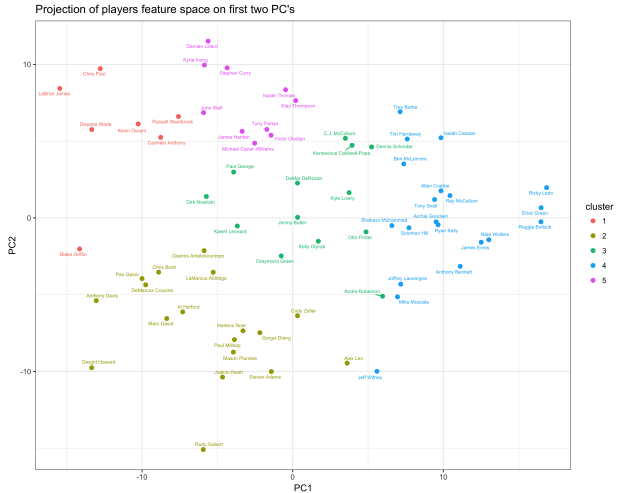


Figure: Projection of Players Feature Space on First Two PC's

Data Analysis

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

Based on the results of clustering, we would like to analyze why player A and player B are in the same cluster.

- James Harden vs. Micheal Cart-Williams
- Paul Millsap vs. Nerlen Noel
- Tony Parker vs. Victor Oladipo

Data Analysis

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

For James Harden vs. Micheal Cart-Williams, We compared their 3 point rate during last four regular seasons.

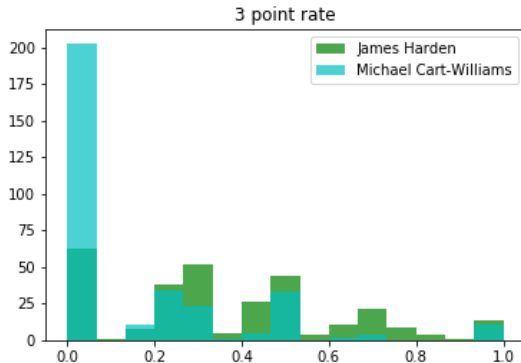


Figure: 3 point rate between James Harden vs. Micheal Cart-Williams in 4 regular season are similar

Data Analysis

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

For Paul Millsap vs. Nerlen Noel, we compared their BLK during last four regular seasons.

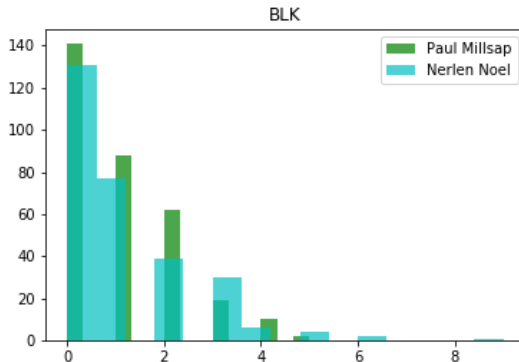


Figure: Plus/Minus between Paul Millsap vs. Nerlen Noel in 4 regular season are similar

Data Analysis

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

For Tony Parker vs. Victor Oladipo, we compared their Field Goal% during last four regular seasons.

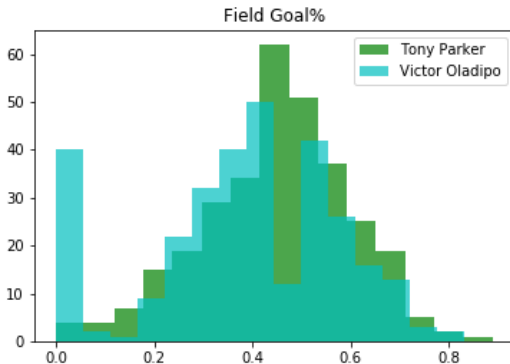


Figure: Field Goal% between Tony Parker vs. Victor Oladipo in 4 regular season are similar

What we expect

The Next
Super Star in
NBA?!

STA 141B

Motivation

Methodology

Data

Ideally
Conclusion

Ideally, we would exploit an automatic clustering method that can correctly split all players into several clusters, and based on the similar pattern between all stars and potential stars to identify a forthcoming super star!

- Micheal Cart-Williams
- Nerlen Noel
- Victor Oladipo