Classifying Modern NBA Players (& Teams)

Tejas Bala

Motivation

Provide a wish list of Free Agent players for NBA teams in the offseason based off of their goals, areas for improvement, salary constraints, roster composition and style of play.

Kemba Walker: I trust Hornets to build appealing roster

Which NBA teams can chase superstar free agents?

Here's our look at who can make big moves in July.

NBA free agents: Team-by-team lists for 2019 and 2020

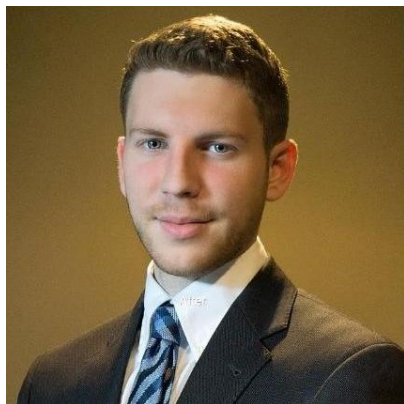Who are the free agents this year and next year? We have the lists.

Top of the East can now swing superstar free agency

Scaling Back Scope

Due to timeline of this project and the large scope of a Recommendation Engine I wanted to scale back to a more feasible goal:

Provide similarity models of Teams and Players to inform a Recommendation Engine.

Domain Expert

Eric Sidewater

- Journalist & Data Scientist
- Heavily involved with publications on Basketball Analytics
- Good relationships with personnel in professional sports industry
- Currently runs Sidewater Analytics

Data Sources

Basketball-Reference

Team per Game Stats

Team Per Game Stats * Playoff teams Share & more ▼ Glossary

Rk	Team	G	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	New Orleans Pelicans	61	240.0	43.4	90.8	.478	10.1	28.9	.348	33.3	61.8	.539	18.9	24.5	.770	11.0	36.0	47.0	26.8	7.2	5.7	14.4	21.5	115.7

Opponent per Game Stats

Opponent Per Game Stats Share & more ▼ Glossary

Rk	Team	G	MP	FG	FGA	FG%	3P	3PA	3P% ▼	2P	2PA	2P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	Washington Wizards	60	243.3	42.9	89.7	.478	12.3	32.7	.377	30.6	57.0	.536	18.9	24.2	.783	11.4	36.2	47.6	26.1	7.8	4.5	16.8	21.6	117.0

Misc. Stats

Miscellaneous Stats Share & more ▼ Glossary

																Offense Four Factors				Defense Four Factors					
Rk	Team	Age	W	L	PW	PL	MOV	SOS	SRS	ORtg	DRtg	NRtg	Pace	FTr	3PAr	TS%	eFG%	TOV%	ORB%	FT/FGA	eFG%	TOV%	DRB%	FT/FGA	Arena	Att
1	Milwaukee Bucks	26.8	45	14	45	14	9.66	-0.59	9.08	113.4	104.0	+9.4	102.8	.247	.416	.584	.552	12.4	20.6	.191	.500	12.0	80.5	.165	Fiserv Forum	52

Team Shooting

Team Shooting Current for games through 2019-02-23. Share & more ▼ Glossary

					% of FGA by Distance					FG% by Distance					2-Pt Field Goals				3-Pt Field Goals			
															Dunks		Layups		Corner		Heaves	
Rk	Team	G	MP	FG%	Dist.	2P	0-3	3-10	10-16	16-3pt	3P	2P	0-3	3-10	10-16	16-3pt	3P	%Ast'd	%FGA	Md.	%FGA	Md.	%Ast'd	%3PA	3P%	Att.	Md.
1	Atlanta Hawks	60	14425	.452	13.3	.608	.345	.134	.077	.052	.392	.520	.640	.344	.370	.395	.348	.540	.082	384	.316	873	.824	.172	.401	4	0

Basketball Reference

- Standardized Stats
- Misc Stats
- Contact Info
- Advanced Statistics

Data Collected

NBA Season 2015-2016 to current season (2018-2019)

- Split Train/Test
  - Train: 2015-2016, 2016-2017, 2017-2018
  - Test: 2018-2019

Players

- 450 Players per season (1800 total)

Teams

- 30 Teams per season (120 total)

Modelling Approach

To generate recommendations:

- Similarity metric via clustering analysis
  - Team-Team
    - ex. Team A is need of a Point Guard. They have a very similar play style to Team B. Identify effective Point Guards on Team B, check if any are FA
  - Player-Player
    - ex. Team A is need of a Backup Center. They have a Star Center who is injury prone. Since Team A's system is designed around their Star Center they want to look for Centers who are similar to their Star Center

- Fit transformations and models to training data and applied to testing data

Player Clustering Analysis

Objective

- Identify alike players for match-up and player-replacement/injury hedging purposes

Player Clustering Analysis

Methodology

- Use LDA to maximize separation within Traditional Positions
- Exclude features that are correlated with Player Success
- K-means Clustering modeling

Results

- 8 new classifications of players:
  - Post-Heavy Big, Offensive Big, Defensive Forward, Pace & Space Forward, 3 & D Wing, Slashing Guard, Combo Guard, Primary Ball-Handler/Distributor

https://public.tableau.com/views/NBAPlayerClassification/Dashboard1?:embed=y&:display_count=yes

Team Clustering Analysis

Objective

- Identify groupings of teams in terms of style of play (pacing, 3pt vs 2pt fg freq etc.
  - Uses Player Clustering to calculate Team Roster Composition
- Not: Identify groupings of teams in terms of goals (Championship competitor, Rebuilding, Mid-Low playoff spot competitor) -- too correlated with Wins

Team Clustering Analysis

Methodology

- Uses Player Clustering to calculate Team Roster Composition
- Include 3-point attempt rate, Free Throw Attempt Rate, Pace
- PCA to reduce dimensionality
- K-means Clustering

Results

- 6 clusters of teams based on roster composition and style of play metrics

https://public.tableau.com/views/NBA_Team_Clusters/Dashboard1?:embed=y&:display_count=yes&publish=yes

Evaluation

- Large struggle point for me
- No existing dataset with team or player similarity metrics or groupings
- Worked with Domain Expert to assess clusterings
  - Looked at instances we expected to be similar
  - Looked at instances that surprised us
    - Inspected features used in model to see why items were clustered as such

Future Work

- Identify concrete method to assess clusters
- Implement recommendations based off of these similarity models
- Scale: Use more historical data to track player and team trends in the NBA

Thanks!

Questions, Suggestions, Concerns?