NBA Players Segment Analysis

by John K. Hancock
jkhancock@gmail.com

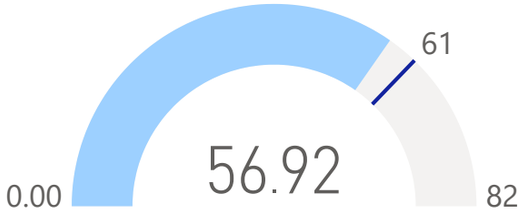
Executive Summary

This report was prepared at the request of San Antonio Spurs owner, James Duarte. The scope of the project was to create two segmentation analyses of NBA players based on their performance in the 2021-22 NBA regular season. The purpose of these segments is to aide in evaluating trade proposals, contract extension negotiations, free agent acquisitions, and even outright release. The analysis successfully identified distinct segments of NBA players. These segments categorize players from marginal up to elite. The report used two unsupervised machine learning algorithms, KMeans Clustering and Gaussian Mixture Models ("GMM") to create and populate the segments. The report structure is as follows:

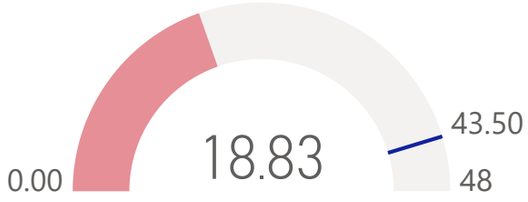
Tab	Description
2.0	A look at league averages for Games and Minutes played. The number of players per positions, and the Age distribution of the players.
3.0	A look at KMeans Clustering Segments
3.1	KMeans Segments Contribution Comparisons
3.2	Comparison of KMean Segments
3.3	Index of Players based on KMeans Segments
4.0	A look at GMM Clustering Segments
4.1	Comparison of GMM Segments
4.2	Index of Players based on GMM Segments
5.0	Summary

League Averages

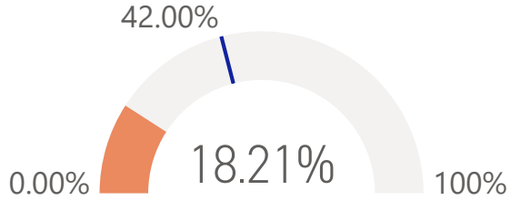
Average Games Played



Average Minutes Played

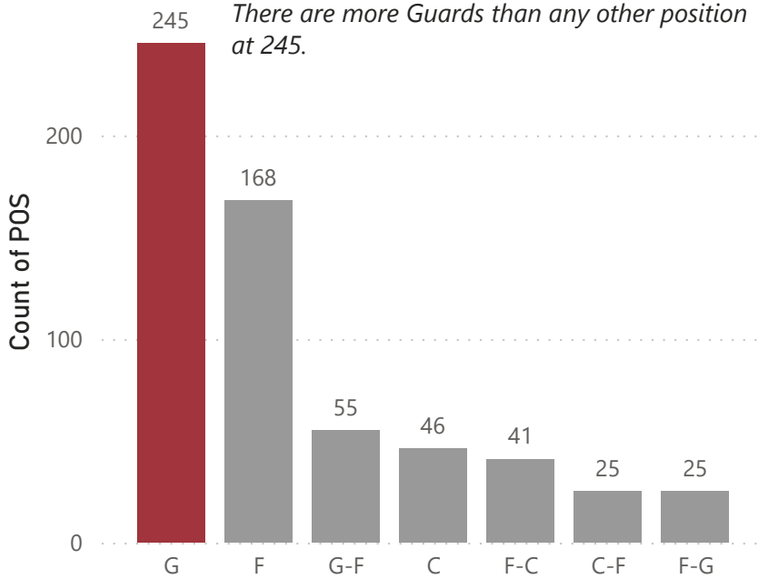


Average Usage %



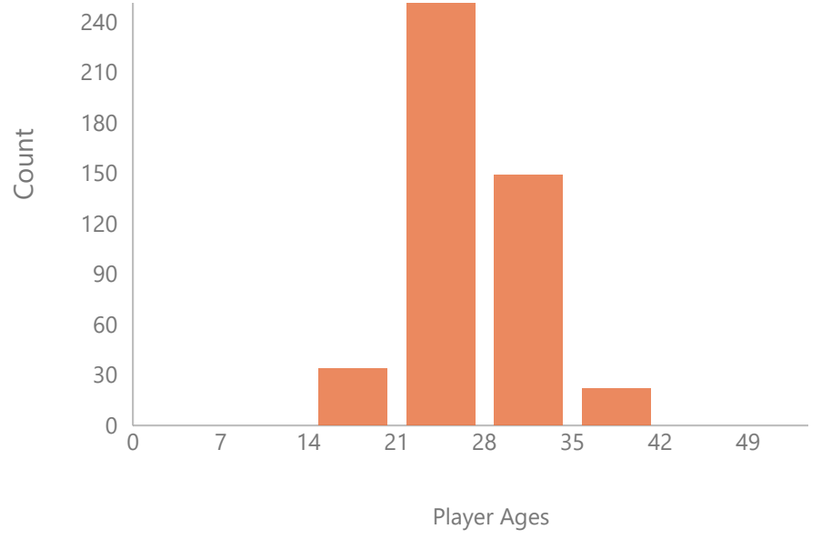
Players per Position

There are more Guards than any other position at 245.



Age Distribution of Players

Players' average age is between 21 and 28 with only 22 players aged 35 and over.

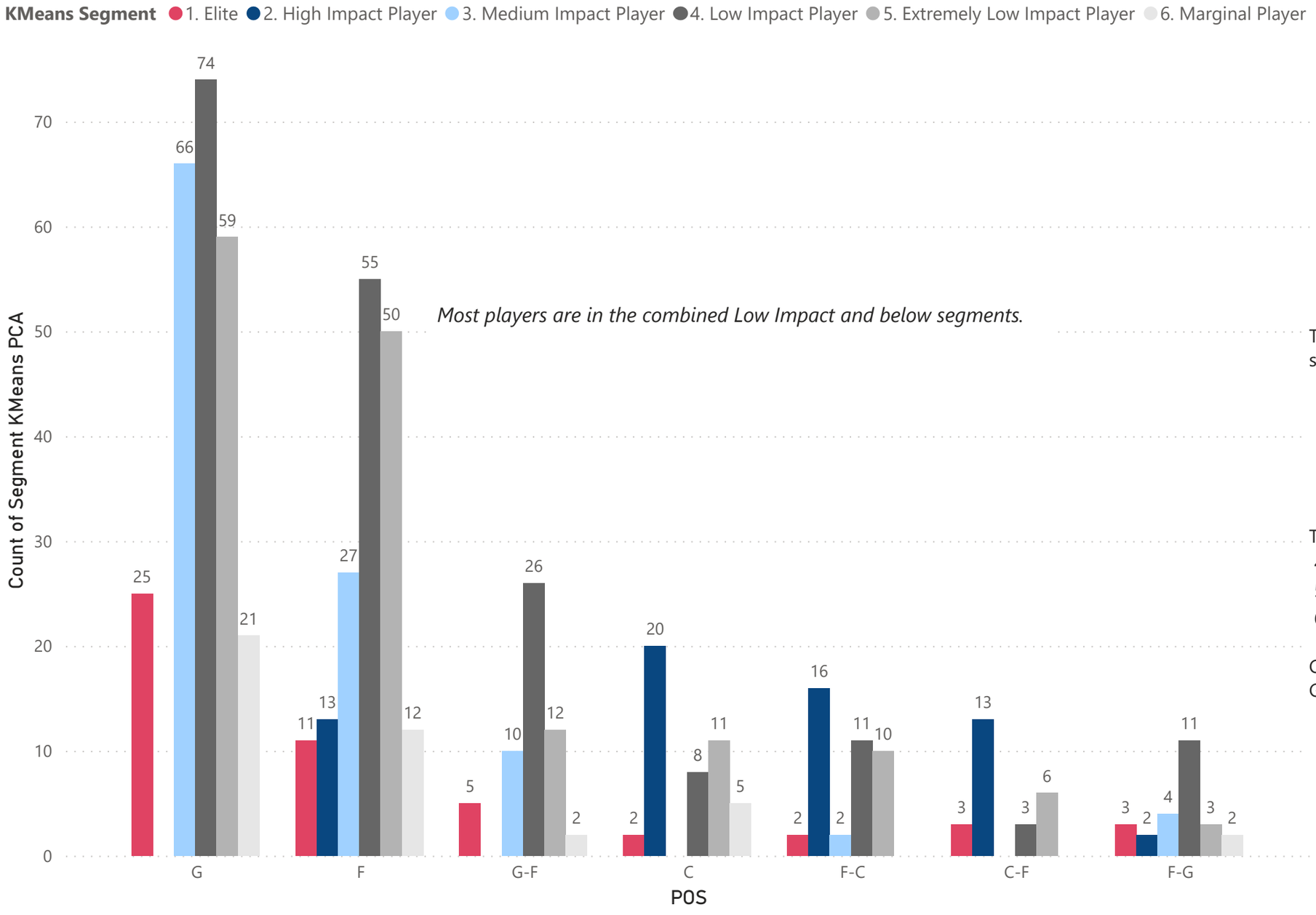


Highest Combination of Games Played and Minutes

Mikal Bridges

FULL NAME	GP	MPG	Usage %	AGE
Deni Avdija	82	24.20	16.30%	21.27
Dwight Powell	82	21.90	14.00%	30.73
Kevon Looney	82	21.10	12.30%	26.18
Mikal Bridges	82	34.80	15.00%	25.61
Saddiq Bey	82	33.00	21.30%	23.01
Alec Burks	81	28.60	17.70%	30.73
Buddy Hield	81	32.10	22.45%	29.31
Patty Mills	81	29.00	16.40%	33.66
Terance Mann	81	28.60	16.10%	25.48
Chris Boucher	80	21.10	17.80%	29.25
Dorian Finney-Smith	80	33.00	13.80%	28.94

Number of Positions per KMeans Segment



Most players are in the combined Low Impact and below segments.

KMeans Clustering Algorithm

The KMeans Clustering Algorithm is an "unsupervised learning algorithm". It looks for patterns in the data using features. The algorithm works by grouping data together into a fixed number of clusters. This number of clusters is the "K" in KMeans clustering. Data points are assigned to clusters based on their distance from the center of the cluster.

The algorithm clustered NBA player data into six segments:

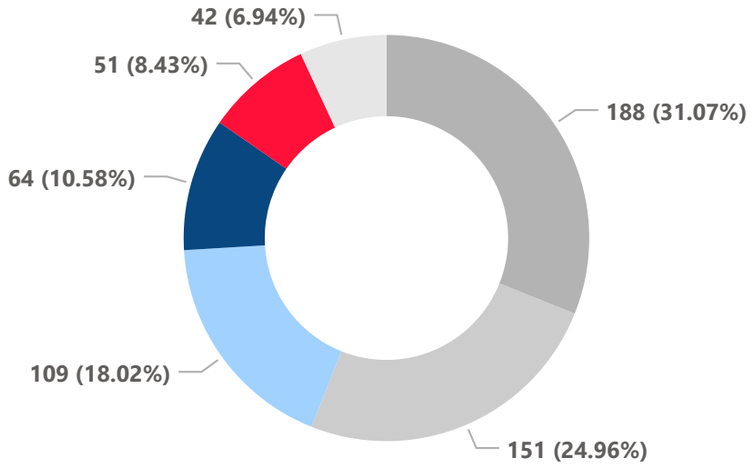
1. **Elite**: The best players in the NBA including the last 10 NBA MVPs
2. **High Impact**: The second tier of the best players
3. **Medium Impact**: These are the average players

The next three segments of players:

4. **Low Impact**
5. **Extremely Low Impact**
6. **Marginal**

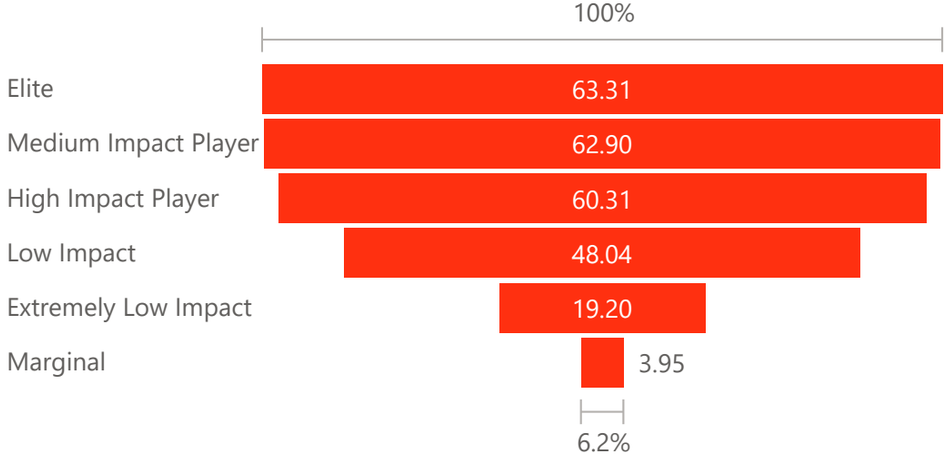
Contain very weak contributors as evidenced by their Games and Minutes played.

KMeans Segments

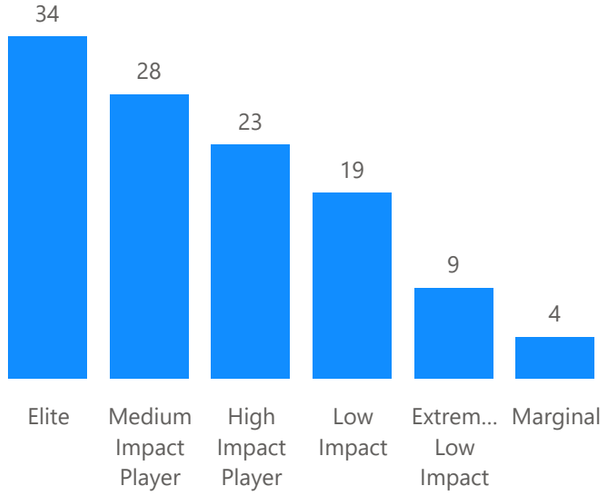


Segment KMeans PCA
- Low Impact
- Extremely Low Impact
- Medium Impact Player
- High Impact Player
- Elite
- Marginal

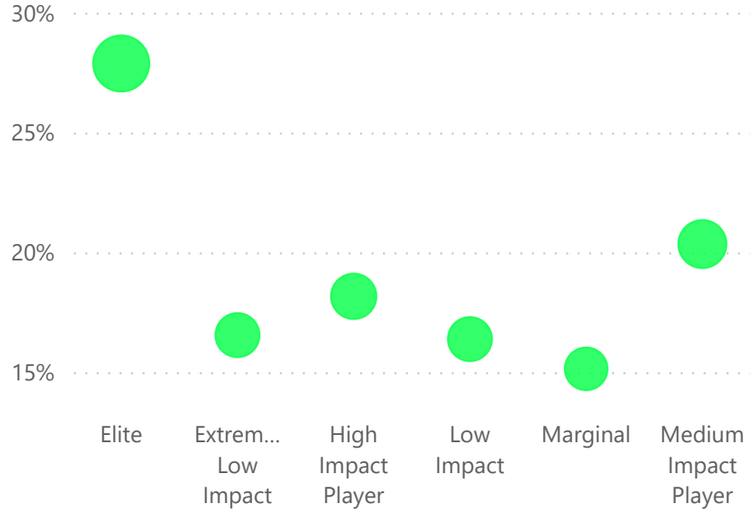
Average Games Played



Average Minutes per Game



Average Usage %

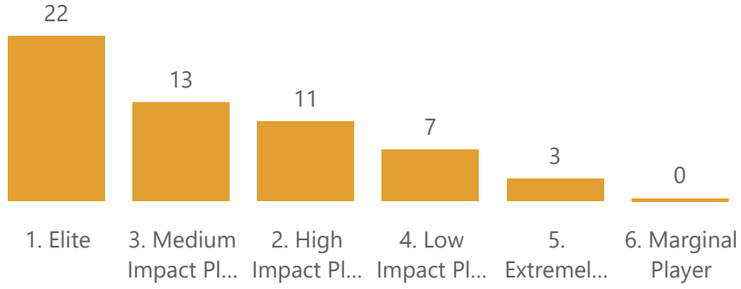


Comparisons of KMeans Segments

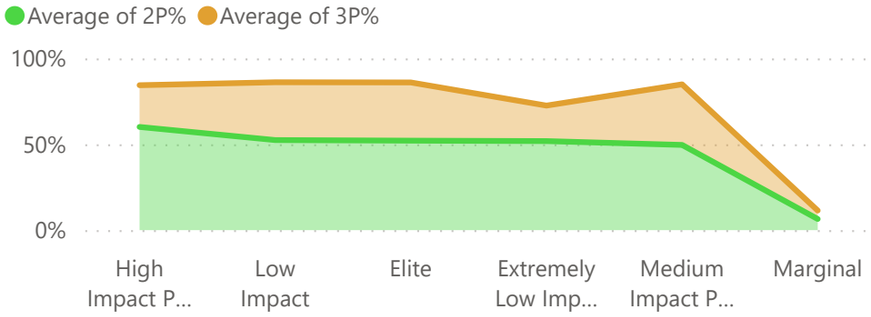
KMeans Segment

- [ ] Select all
- [ ] 1. Elite
- [ ] 2. High Impact Player
- [ ] 3. Medium Impact Player
- [ ] 4. Low Impact Player
- [ ] 5. Extremely Low Impact Player
- [ ] 6. Marginal Player

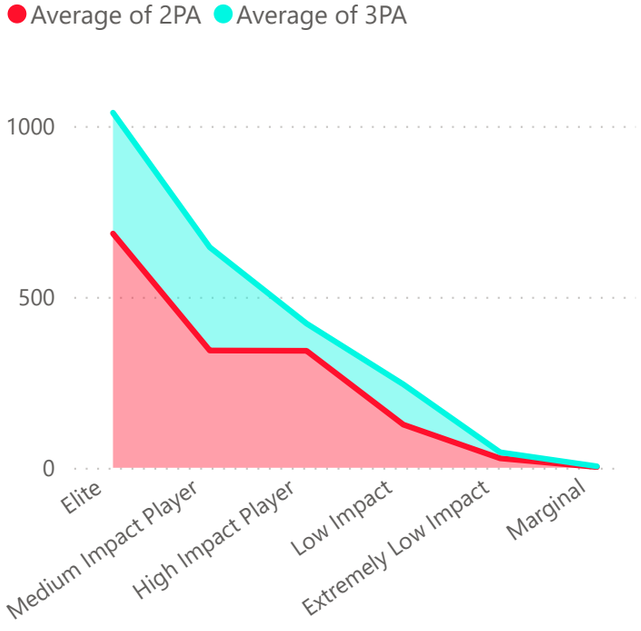
Points per Game



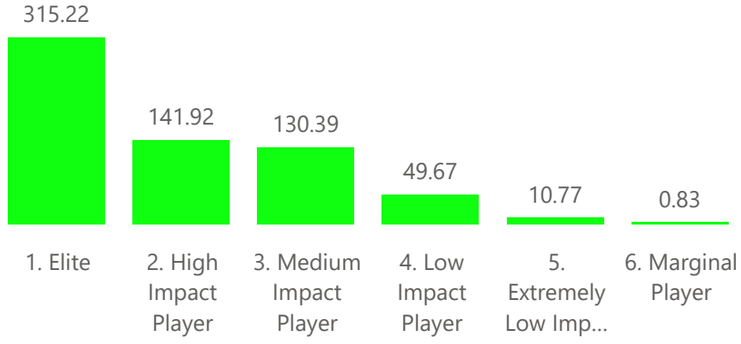
Average Two and Three Points Made %



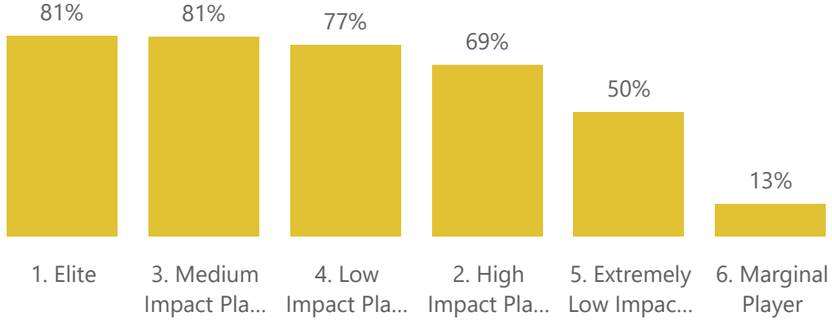
Average Two and Three Points Attempts



Free Thow Attempts



Percentage of Free Throws Made



Rebounds per Game

KMeans Segment

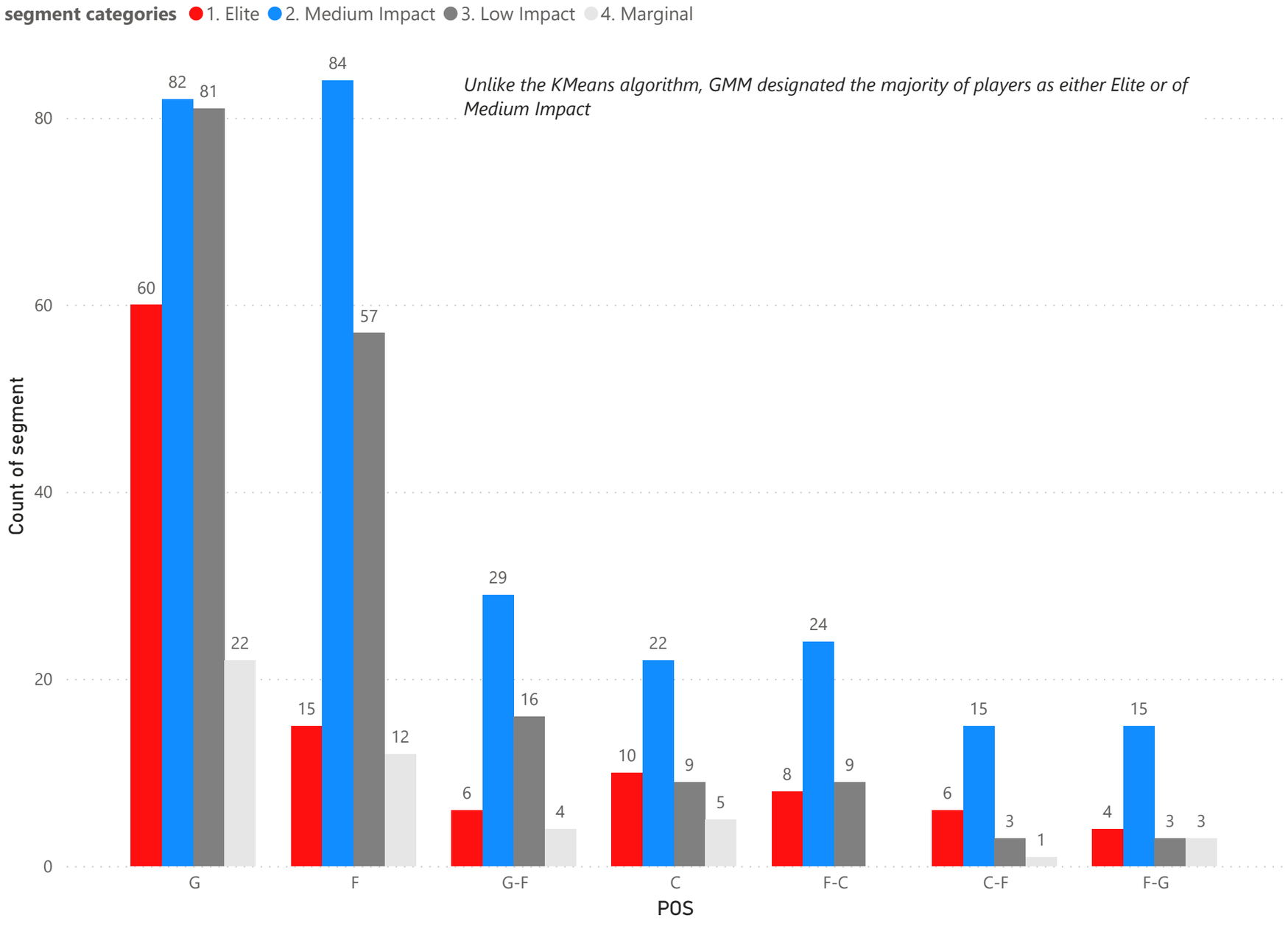
- [ ] 1. Elite
- [ ] 2. High Impact Player
- [ ] 3. Medium Impact Player
- [ ] 4. Low Impact Player
- [ ] 5. Extremely Low Impact Player
- [ ] 6. Marginal Player

FULL NAME

- [ ] Aaron Gordon
- [ ] Aaron Henry
- [ ] Aaron Holiday
- [ ] Aaron Nesmith
- [ ] Aaron Wiggins
- [ ] Abdel Nader
- [ ] Ade Murkey

Segment KMeans PCA	FULL NAME	AGE	Games Played	Minutes per Game	Assists per game	Points per Game	Rebounds per Game	Blocks per Gamee	Steals per Game	2PA	2P%	3PA	3P%	FTA	FT%
Elite	Anthony Edwards	20.68	72	34.30	3.80	21.30	4.80	0.64	1.46	642	52%	602.00	36%	280	79%
Elite	Bam Adebayo	24.73	56	32.60	3.40	19.10	10.10	0.79	1.43	723	56%	6.00	0%	340	75%
Elite	Bradley Beal	28.79	40	36.00	6.60	23.20	4.70	0.38	0.90	561	51%	210.00	30%	203	83%
Elite	Brandon Ingram	24.61	55	34.00	5.60	22.70	5.80	0.45	0.62	757	50%	226.00	33%	322	83%
Elite	Cade Cunningham	20.54	64	32.60	5.60	17.40	5.50	0.67	1.22	668	47%	363.00	31%	168	85%
Elite	Chris Paul	36.93	65	32.90	10.80	14.70	4.40	0.31	1.86	537	56%	199.00	32%	202	84%
Elite	Cole Anthony	21.90	65	31.70	5.70	16.30	5.40	0.26	0.71	521	43%	391.00	34%	253	85%
Elite	Damian Lillard	31.74	29	36.40	7.30	24.00	4.10	0.38	0.59	268	49%	284.00	32%	181	88%
Elite	D'Angelo Russell	26.13	65	31.90	7.10	18.10	3.30	0.34	0.95	457	49%	518.00	34%	240	83%
Elite	Darius Garland	22.21	68	35.70	8.60	21.70	3.30	0.10	1.31	719	51%	454.00	38%	241	89%
Elite	De'Aaron Fox	24.31	59	35.30	5.60	23.20	3.90	0.42	1.15	843	52%	249.00	30%	348	75%
Elite	Dejounte Murray	25.56	68	34.80	9.20	21.10	8.30	0.34	2.04	947	50%	294.00	33%	247	79%
Elite	DeMar DeRozan	32.68	76	36.10	4.90	27.90	5.20	0.32	0.89	1393	52%	142.00	35%	593	88%
Elite	Devin Booker	25.44	68	34.50	4.80	26.80	5.00	0.38	1.13	943	51%	478.00	38%	363	87%
Elite	Domantas Sabonis	25.94	62	34.15	5.40	18.90	12.20	0.40	0.93	324	61%	62.00	28%	166	74%
Elite	Donovan Mitchell	25.59	67	33.80	5.30	25.90	4.20	0.18	1.48	722	53%	654.00	36%	313	85%
Elite	Draymond Green	32.10	46	28.90	7.00	7.50	7.30	1.09	1.35	203	59%	54.00	30%	91	66%
Elite	Fred VanVleet	28.12	65	37.90	6.70	20.30	4.50	0.55	1.71	455	44%	642.00	38%	222	87%
Elite	Giannis Antetokounmpo	27.35	67	32.90	5.80	29.90	11.60	1.36	1.07	1003	62%	242.00	29%	766	72%
Elite	Ja Morant	22.67	57	33.10	6.70	27.40	5.70	0.39	1.16	920	54%	256.00	34%	415	76%

Number of Positions per Gaussian Mixture Model Segment



Gaussian Mixture Model Clustering Algorithm

The second unsupervised learning algorithm, Gaussian Mixture Model ("GMM") looks for a mixture of multi-dimensional Gaussian probability distributions to fit the data. It's more of a probability distribution or a mixture of different distributions.

GMM clustered the data among just four segments.

1. **Elite**: The best players in the NBA including the last 10 NBA MVPs
2. **Medium Impact**: The second tier of the best players

The next two segments of players:

3. **Low Impact**
4. **Marginal**

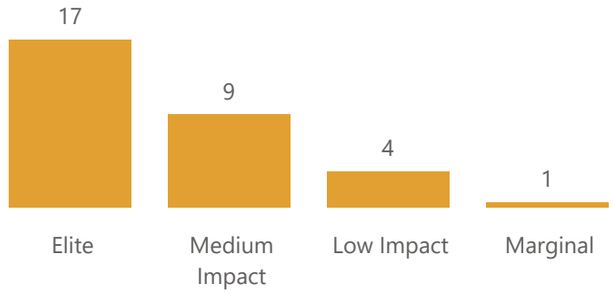
Are for players who do not play in many games nor get significant minutes. These are very low contribution players.

Comparisons of GMM Segments

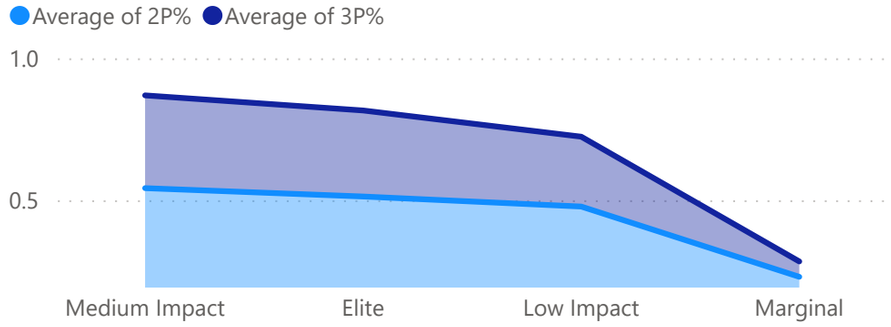
GMM Segment

- [ ] Select all
- [ ] 1. Elite
- [ ] 2. Medium Impact
- [ ] 3. Low Impact
- [ ] 4. Marginal

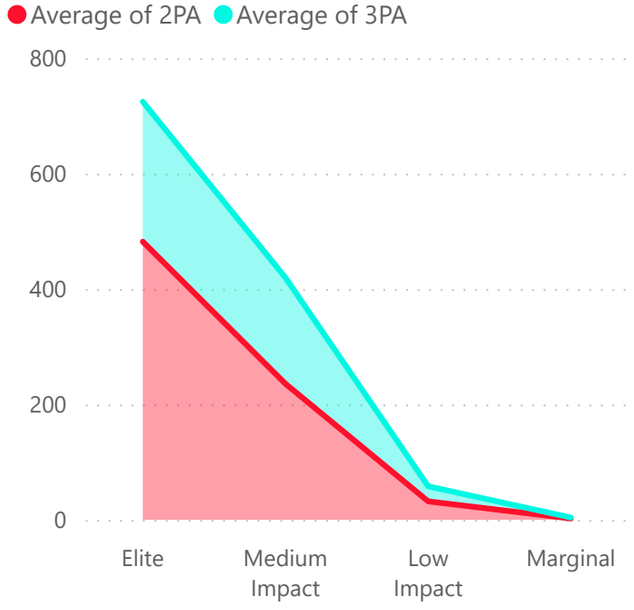
Points per Game



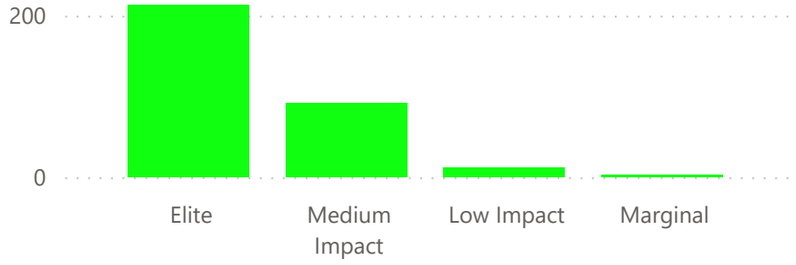
Average Two and Three Points Made %



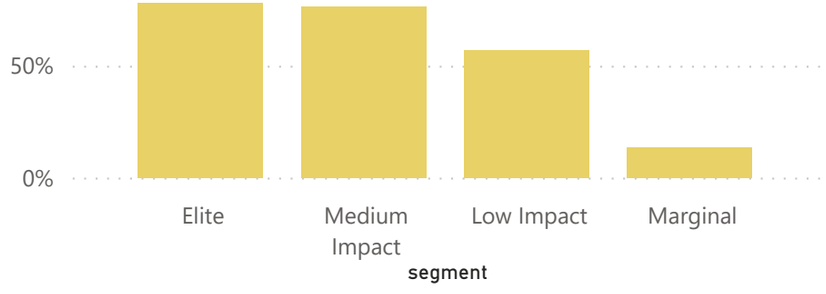
Average Two and Three Points Attempts



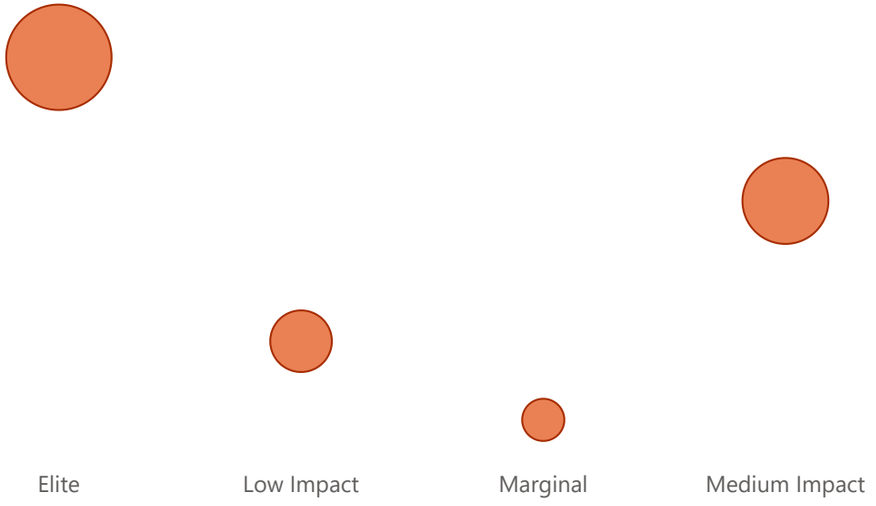
Free Throw Attempt by Segment



Free Throw% by Segment



Rebounds per Game



segment categories

- [ ] 1. Elite
- [ ] 2. Medium Impact
- [ ] 3. Low Impact
- [ ] 4. Marginal

FULL NAME

- [ ] Aaron Gordon
- [ ] Aaron Henry
- [ ] Aaron Holiday
- [ ] Aaron Nesmith
- [ ] Aaron Wiggins
- [ ] Abdel Nader
- [ ] Ade Murkey

segment	FULL NAME	AGE	Games Played	Minutes per Game	Assists per game	Points per Game	Rebounds per Game	Blocks per Gamee	Steals per Game	2PA	2P%	3PA	3P%	FTA	FT%
Elite	Alex Caruso	28.12	41	28.00	4.00	7.40	3.60	0.37	1.73	130	46%	126.00	33%	73	80%
Elite	Andre Drummond	28.67	73	20.35	1.70	8.95	9.55	0.96	1.02	215	58%	2.00	0%	82	52%
Elite	Andre Iguodala	38.20	31	19.40	3.70	4.00	3.20	0.71	0.87	47	62%	74.00	23%	20	75%
Elite	Anfernee Simons	22.84	57	29.50	3.90	17.30	2.60	0.11	0.53	355	49%	442.00	41%	116	89%
Elite	Anthony Davis	29.09	40	35.10	3.10	23.20	9.90	2.25	1.23	625	57%	70.00	19%	244	71%
Elite	Anthony Edwards	20.68	72	34.30	3.80	21.30	4.80	0.64	1.46	642	52%	602.00	36%	280	79%
Elite	Bam Adebayo	24.73	56	32.60	3.40	19.10	10.10	0.79	1.43	723	56%	6.00	0%	340	75%
Elite	Brad Wanamaker	32.71	23	20.20	4.60	5.25	2.80	0.14	1.12	30	40%	8.00	12%	12	95%
Elite	Bradley Beal	28.79	40	36.00	6.60	23.20	4.70	0.38	0.90	561	51%	210.00	30%	203	83%
Elite	Brandon Ingram	24.61	55	34.00	5.60	22.70	5.80	0.45	0.62	757	50%	226.00	33%	322	83%
Elite	Brandon Williams	22.38	24	26.70	3.90	12.90	3.10	0.42	1.00	176	42%	106.00	29%	97	70%
Elite	Cade Cunningham	20.54	64	32.60	5.60	17.40	5.50	0.67	1.22	668	47%	363.00	31%	168	85%
Elite	Cameron Payne	27.67	58	22.00	4.90	10.80	3.00	0.29	0.69	376	45%	211.00	34%	89	84%
Elite	Caris LeVert	27.63	58	30.45	4.15	16.15	3.60	0.36	0.87	299	49%	128.00	32%	96	75%
Elite	Chris Paul	36.93	65	32.90	10.80	14.70	4.40	0.31	1.86	537	56%	199.00	32%	202	84%
Elite	Christian Wood	26.54	68	30.80	2.30	17.90	10.10	0.97	0.78	542	57%	336.00	39%	332	62%
Elite	CJ McCollum	30.56	62	34.50	5.15	22.40	4.40	0.31	1.16	348	51%	234.00	39%	85	69%
Elite	Clint Capela	27.90	74	27.60	1.20	11.10	11.90	1.26	0.76	603	61%	1.00	0%	169	47%
Elite	Cole Anthony	21.90	65	31.70	5.70	16.30	5.40	0.26	0.71	521	43%	391.00	34%	253	85%
Elite	Damian Lillard	31.74	29	36.40	7.30	24.00	4.10	0.38	0.59	268	49%	284.00	32%	181	88%

Summary

These models provide insightful data analysis to aide decision-making. The two unsupervised machine learning algorithms created a different number of segments, KMeans (6) and GMM (4). GMM had more than double the number of Elite players. Both algorithms had similar numbers for Low Impact and Marginal players. Games Played and Minutes per Game are significant features in determining these segments. GMM also aggregated the High Impact and Medium Impact into just one segment, Medium Impact. This is surprising as the GMM algorithm is more flexible in terms of assigning data to clusters.

The best way to use this report would be to check the KMeans and GMM indexes (Tabs 3.3 and 4.2) for a player's segment. For a more detailed report showing the code that created the segments and the process that was undertaken, please visit this link: **A Comparison of Unsupervised Machine Learning Algorithms, KMeans Clustering and Gaussian Mixture Model**