Deciphering NBA Game 'Personality' Clusters:

An Exploration of Success Factors

CASSIS

Abstract

"This is going to be a slugfest," remarked Nick Nurse, anticipating the Raptors' 2022 Eastern Conference Playoff duel with the 76'ers [1]. Nurse's statement underscores a conventional belief, be it a shootout or a track meet, each NBA game unveils a unique personality. In this research, through the lens of k-means clustering and frequent itemset analysis, we delve into advanced box scores from NBA games, spotlighting these personality traits and pinpointing the critical factors for success in each game type.

Phase 1

We begin by collecting box scores from the modern era (2014-2023). The box scores are aggregated to create team-level summaries of various statistics, which will be used to build 'personality' descriptions of each game.

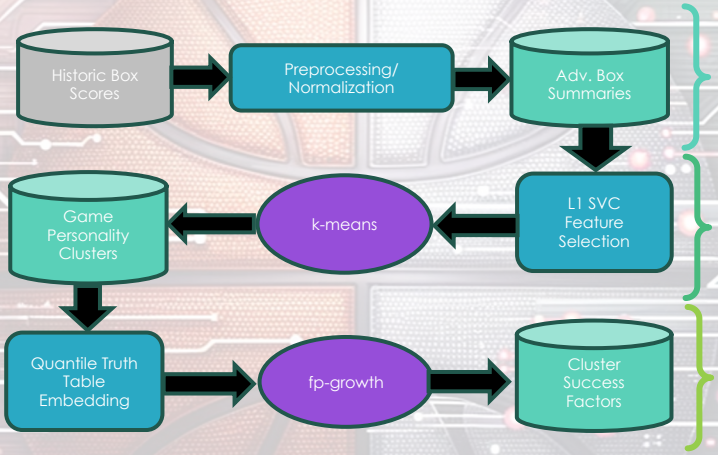
Phase 2

Feature selection is then implemented to narrow the scope for the clustering process. Elbow and silhouette plots were used to target a cluster count of 5. Principal Component Analysis was used to visualize the clusters and their approximate feature distribution and Information Gain to identify/rank winning features.

Phase 3

Once clusters had been identified, representative sets of 150 were extracted from each cluster, and quantile statistics were generated by column. Each game row was then converted into a series of truth values indicating whether the statistic eclipsed the quantile threshold set by that cluster's teams. The truth table was converted to a series of itemsets and fed to the pattern frequency algorithm, which upon pruning, provided our multi-variate success strategies.

'Personality' Cluster Pipeline



Results

The 5 resulting clusters, with an average distortion of 20,715.7 across 1,468 samples, show clear distinctions in box score statistics. Clusters 2 and 5 stand out as potential 'Shootout' and 'Slugfest' game types.

Using information gain, the 4 most win-predictive features were identified for each cluster. While field goals made (approximating score) was key in all clusters, a diverse set of 12 total features emerged as success factors.

Multi-variate rules analysis revealed higher win probability linked to basketball strategies like shooting efficiently and dominating the offensive glass. Analyzing feature combinations exposed more nuanced strategic insights.

fpgrowth Example Rules

Narrowed to cluster #4, each rule represents an elite winning strategy, seen in only 1-1.55% of teams but with the highest impact on success

- 80th percentile: 3-point percentage, field goal percentage
- 70th percentile: free throw percentage, personal fouls

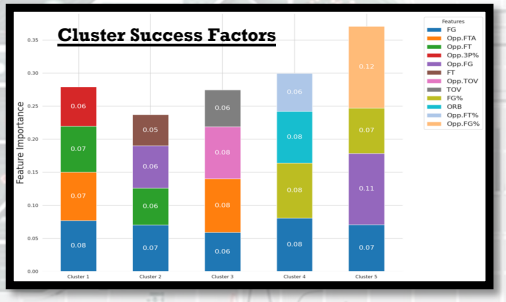
- 90th percentile: 3-point percentage
- 80th percentile: free throw percentage
- 70th percentile: field goal percentage

- 80th percentile: free throw percentage, field goals made
- 70th percentile: field goal percentage, personal fouls

- 80th percentile: free throws made, field goals made
- 70th percentile: field goal percentage

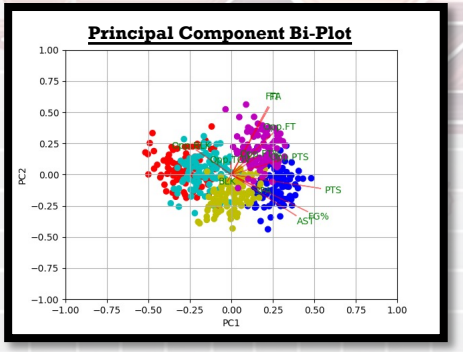
- 90th percentile: offensive rebounds
- 70th percentile: field goal percentage, field goals made

Produce a **2x** likelihood of winning



Stat-Category Performance by Cluster

PTS	FG	FG%	3P	3P%	FT	FT%	ORB	TRB	AST	STOCK	TOV	PF	Win%	#
112.52	39.60	46.12%	9.61	35.07%	23.71	77.63%	11.13	44.77	22.90	12.99	14.27	24.05	51.3%	1
122.83	45.39	50.76%	13.89	42.90%	18.15	78.58%	10.00	42.65	28.23	12.22	12.56	19.76	67.3%	2
99.10	37.48	43.96%	8.19	31.06%	15.95	74.83%	10.52	41.05	21.42	11.50	13.31	18.33	3.33%	3
111.45	42.41	49.44%	11.83	41.60%	14.79	76.64%	9.59	46.43	26.92	13.56	13.21	18.65	97.3%	4
95.21	35.05	42.33%	7.34	30.18%	17.76	76.09%	10.80	46.34	19.71	12.99	13.54	19.34	56.0%	5
OPPONENT														#
110.88	39.10	45.55%	9.77	34.94%	22.89	78.01%	10.61	43.59	22.35	12.97	14.29	24.94	48.7%	1
118.60	44.32	49.62%	12.44	39.70%	17.52	76.99%	10.09	41.82	26.21	11.55	13.08	20.22	32.7%	2
110.74	42.15	49.70%	11.28	40.36%	15.16	77.46%	9.55	44.08	25.21	13.01	12.99	19.31	96.7%	3
96.11	36.06	41.89%	8.50	31.06%	15.50	75.11%	10.30	41.76	20.83	11.39	13.73	18.20	2.7%	4
92.06	34.38	41.17%	7.52	30.38%	15.77	75.15%	10.48	44.66	19.34	12.56	13.76	20.62	44.0%	5



Conclusion

Cluster analysis confirms that NBA games exhibit distinct personalities, characterized by factors like passing, scoring, shooting efficiency, and physicality. The success factors and association rules for each cluster reveal diverse winning dynamics across game types.

Future research could explore these distinctions further, identifying optimal player archetypes and play-actions per cluster. Enhancements to rule generation and pruning could yield a more scalable analytical pipeline.

In summary, unsupervised learning powerfully illuminates the complex factors shaping winning strategies in each game type, offering a competitive edge through data-driven insights.

Koi Stephanos
dkstephanos@gmail.com