



Licenciatura em Ciência de Dados - 3º ano

Clusters de performance na NBA

Projeto Aplicado em Ciência de Dados II

14 de dezembro de 2024

Discentes: João Dias nº 110305 / Felipe Pereira nº 110861 / David Franco nº 110733
António Teotónio nº 111283/ João Batista nº 111611

ÍNDICE

RESUMO	2
1. INTRODUÇÃO	3
2. FRAMEWORK METODOLÓGICO	3
3. BUSINESS UNDERSTANDING	3
i. NBA - O jogo & a história	4
ii. Objetivo do projeto	4
iii. Trabalhos semelhantes	5
4. DATA UNDERSTANDING	6
i. Initial Data	6
ii. Data Quality	8
iii. Data Exploration	8
5. DATA PREPARATION	11
i. Valores omissos	11
ii. Feature engineering e transformação de variáveis	11
6. MODELING	12
i. Análise de componentes principais (PCA)	13
ii. k-Means Clustering	15
7. EVALUATION	16
8. CONCLUSÃO	24
9. LIMITAÇÕES E TRABALHOS FUTUROS	24
BIBLIOGRAFIA	25

RESUMO

Este projeto teve como objetivo analisar a performance dos jogadores da NBA na temporada regular 2023-2024, utilizando técnicas de Análise de Componentes Principais e clustering com k-means para identificar padrões e agrupar jogadores com características semelhantes. A análise visou não só avaliar o desempenho individual dos jogadores, mas também entender como diferentes perfis de jogadores contribuem para o sucesso das equipas ao longo da época.

Os dados utilizados para a análise foram obtidos com recurso a web scraping do website Basketball Reference, que fornece estatísticas detalhadas dos jogadores. A PCA foi aplicada para reduzir a dimensionalidade dos dados e identificar as principais variáveis que afetam o desempenho dos jogadores. A primeira componente (PC1) relacionou-se com a produção ofensiva, enquanto a PC2 foi associada ao desempenho defensivo e aos ressaltos. A PC3 refletiu o impacto dos jogos em casa, e a PC4 indicou a contribuição geral para o sucesso da equipa. A PC5 mostrou a eficiência ofensiva, especialmente em arremessos de três pontos e assistências.

Com estas componentes, aplicou-se o k-means com $k=5$, resultando em clusters que representam diferentes perfis de jogadores. O Cluster 0 é composto por jogadores menos ativos, frequentemente em início ou fim de carreira, com baixa contribuição ofensiva e defensiva. O Cluster 1 inclui jogadores de papel secundário e de banco, com um desempenho moderado tanto ofensivamente quanto defensivamente. O Cluster 2 reúne os jogadores estrela, que são altamente ativos e se destacam em ambas as extremidades do jogo, ofensiva e defensivamente. O Cluster 3 agrupa jogadores defensivos, que se destacam na defesa, principalmente em ressaltos, com um desempenho ofensivo equilibrado. Por fim, o Cluster 4 reúne *role players* ofensivos, com bom desempenho em pontos e percentagens de lançamentos.

Para avaliar a performance global dos jogadores em cada um dos clusters, criou-se uma métrica personalizada, que combina várias estatísticas individuais, ponderadas de acordo com a sua relevância. Esta fórmula calcula a eficiência de um jogador em função dos pontos (PTS), assistências (AST), ressaltos (TRB), bloqueios (BLK), roubos de bola (STL) e turnovers (TOV), normalizando pela quantidade de minutos jogados (MP).

A análise desta métrica, juntamente com a comparação com os NBA Awards da época 2023/24, revelou que fatores como o impacto no jogo, o papel dentro da equipa e elementos intangíveis, como liderança e consistência ao longo da temporada, são cruciais para avaliar o desempenho dos jogadores. A crescente relevância dos jogadores de banco (bench players) também se destacou, pois estes desempenham um papel essencial no sucesso coletivo das equipas, equilibrando as contribuições ofensivas e defensivas. Além disso, foi possível concluir que a atribuição das NBA Awards, além das estatísticas, sendo também influenciada por aspectos intangíveis.

1. INTRODUÇÃO

A NBA é uma liga com uma grande competitividade desportiva e uma vasta disponibilidade de dados estatísticos. Desta forma, é possível realizar análises de desempenho dos jogadores. Com este projeto procurar-se-á estudar a performance dos atletas da NBA ao longo da temporada regular 2023/24, identificando padrões e agrupando-os com base nas suas características e estilos de jogo.

Com recurso ao CRISP-DM, o trabalho será estruturado em etapas claras, desde web scraping dos dados do website Basketball Reference para a aquisição dos *datasets* iniciais até à análise e avaliação dos resultados. Através de métodos como a análise de componentes principais e clustering, espera-se identificar perfis de jogadores, de forma a perceber as características, desempenho e as diferenças entre os atletas, como também insights importantes que marcaram a última temporada e que permitem fazer uma caracterização dos atletas.

2. FRAMEWORK METODOLÓGICO

Para o presente relatório, adotar-se-á o framework metodológico do CRISP-DM (Cross-Industry Standard Process for Data Mining), uma das principais metodologias em Ciência de Dados, que através da sua abordagem sistemática composta pelas seis fases, que se podem ver na Fig. 1, permite uma organização estruturada e eficaz do projeto.

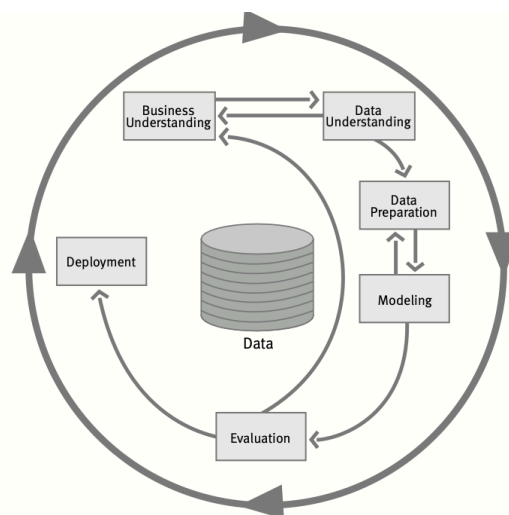


Figura 1 - Framework geral do CRISP-DM

3. BUSINESS UNDERSTANDING

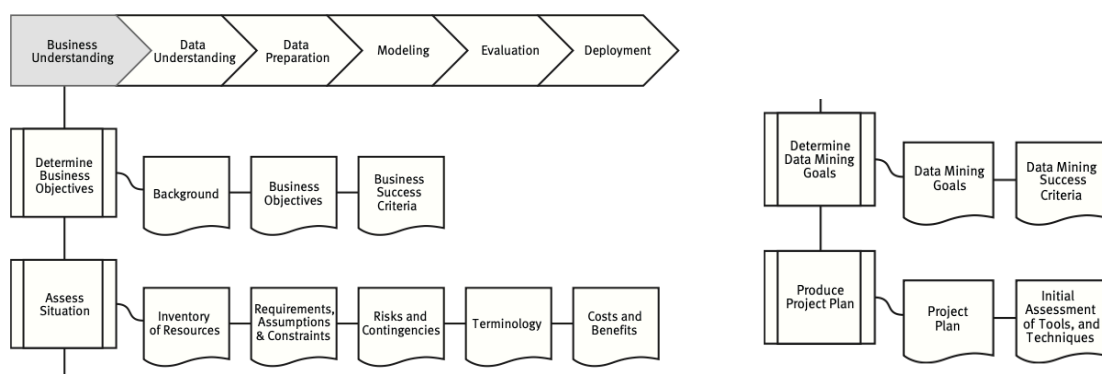


Figura 2 - Descrição das tarefas associadas à fase de Business Understanding

A primeira fase de Business Understanding, para a compreensão do problema, pode ser descrita na Fig. 2, e será estruturada em 3 grandes secções. Por um lado procurar-se-á entender bem como funciona a NBA e descrever em linhas gerais alguns marcos históricos importantes e relevantes para este projeto. De seguida, são definidos os objetivos concretos do trabalho. E, por fim, analisar-se-á outros trabalhos relevantes que serviram de inspiração.

i. NBA - O jogo & a história

A época regular da NBA decorre de outubro a abril, durante a qual cada equipa joga 82 jogos, sendo 41 em casa e 41 fora. As 30 equipas estão divididas em duas conferências, a Este e a Oeste. No final da época regular, as oito melhores equipas de cada conferência, com base no número de vitórias, avançam para os playoffs. Desde 2020, existe um torneio de play-in, onde as seis primeiras equipas de cada conferência passam diretamente e as equipas classificadas entre o 7.º e o 10.º lugar de cada conferência competem pelas duas últimas vagas nos playoffs. Os playoffs são uma competição de eliminação direta disputada em séries à melhor de sete jogos. Por outras palavras, a primeira equipa a vencer quatro jogos avança para a próxima fase.

A medição de estatísticas na NBA começou na época de 1946-1947, o ano de fundação da liga, com registo de dados básicos como pontos, ressaltos e assistências. A partir dos anos 70 e 80, surgiram novas estatísticas, como roubos de bola, bloqueios e percentagens de lançamento. Nos anos 2000, a introdução de tecnologias avançadas permitiu o desenvolvimento de métricas mais sofisticadas.

As estatísticas desempenham um papel fundamental no basquetebol moderno. São utilizadas por treinadores para ajustar estratégias, analisar adversários e medir o desempenho dos jogadores em tempo real. Em scouting, ajudam a avaliar o potencial de atletas para draft, trocas ou contratações. Além disso, as métricas avançadas, como a eficiência ofensiva e defensiva, permitem que as equipas desenvolvam abordagens táticas baseadas em dados. Para os fãs e os meios de comunicação, enriquecem a experiência ao fornecerem informações detalhadas e narrativas que contextualizam o jogo. Em suma, são indispensáveis tanto para decisões em campo como na gestão global das equipas.

ii. Objetivo do projeto

O objetivo deste projeto é analisar e explorar a performance dos jogadores da NBA ao longo da época 2023-2024, utilizando métodos que nos permitam identificar padrões e comportamentos distintos entre os jogadores. A análise foca-se em compreender como os jogadores podem ser agrupados em categorias consoante as suas características e avaliar a progressão dos jogadores mais representativos de cada grupo ao longo da época.

Para avaliar a performance global dos jogadores, procurar-se-á definir e calcular uma métrica a partir das estatísticas dos jogadores considerando o desempenho ofensivo, defensivo, entre outros. Esta métrica será fundamental para a análise detalhada dos jogadores e ajudará a identificar os que mais se destacaram ao longo da época.

O objetivo passa por agrupar os jogadores em clusters com características semelhantes. Esses clusters terão como propósito organizar os jogadores em grupos de características similares, facilitando assim a análise de padrões dentro de cada cluster e permitindo uma visão mais clara de diferentes perfis de jogadores. Esta análise dentro de cada cluster irá permitir uma melhor compreensão das características que diferenciam os jogadores e destacar os principais atletas de cada grupo.

iii. Trabalhos semelhantes

No âmbito da análise de desempenho na NBA, têm sido desenvolvidos diversos estudos que procuram identificar padrões de jogo, segmentar jogadores e apoiar a tomada de decisões estratégicas. Estas investigações utilizam abordagens como a análise estatística tradicional e métodos avançados de machine learning, com destaque para o clustering, que permite agrupar atletas com base em características específicas e métricas de desempenho.

Zhang et al. (2020) realizaram um estudo que utilizou um modelo de clustering para agrupar jogadores da NBA com base em atributos antropométricos e experiência de jogo. A análise de 699 jogos equilibrados da época regular permitiu identificar cinco perfis distintos: jogadores com elevada altura e peso, mas baixa experiência (TopHW-LowE); altura e peso médios com experiência média (MiddleHW-MiddleE) ou elevada (MiddleHW-TopE); e altura e peso baixos com baixa (LowHW-LowE) ou média experiência (LowHW-MiddleE). Os resultados destacaram que jogadores de equipas mais fracas eram predominantemente classificados no grupo LowHW-LowE, enquanto os de equipas mais fortes pertenciam ao grupo LowHW-MiddleE, e aqueles que participaram nas finais estavam principalmente no grupo MiddleHW-MiddleE.

Patel (2017) procurou reclassificar os jogadores da NBA com base no seu desempenho durante a época regular de 2016-2017. Utilizando técnicas de redução de dimensionalidade, como t-SNE e análise de componentes principais (PCA), o estudo analisou 18 métricas clássicas de desempenho, reduzindo-as a duas dimensões para visualização. Através do método de clustering k-means, foram identificados quatro grupos de jogadores com estilos de jogo semelhantes. Os resultados indicaram que, enquanto a representação dos jogadores nos clusters era semelhante entre as equipas, as equipas mais fortes tinham jogadores mais distantes dos centros dos clusters no gráfico, sugerindo que o sucesso dessas equipas não se devia apenas a habilidades fundamentais, mas também a fatores avançados ou intangíveis que complementam o desempenho dos atletas.

Richardson (2019) utilizou a análise de componentes principais (PCA) e clustering hierárquico para classificar jogadores da NBA desde a temporada 2000-2001, observando as transições dessas classificações ao longo do tempo. O estudo usou atributos dos jogadores, como distribuições de tipo e localização dos arremessos, assistências, ressaltos, roubos de bola e outras estatísticas ofensivas e defensivas, para identificar cinco componentes principais de estilos de jogo. A análise revelou 12 classificações únicas de jogadores e forneceu insights sobre como os atletas evoluíram ao longo das suas carreiras. Jogadores em início de carreira tendem a transitar para classificações de maior habilidade, enquanto as estrelas da liga começam suas carreiras em classificações de alta habilidade e são mais estáveis dentro delas. As classificações dos *rookies*, por outro lado, mostraram-se menos estáveis, indicando uma maior variabilidade nas suas trajetórias iniciais de desenvolvimento.

Muniz e Flamand (2022) propuseram uma abordagem para agrupar jogadores da NBA com base em dados históricos utilizando uma técnica de detecção de comunidades em grafos de similaridade. O estudo começou com clustering k-means e, em seguida, usou a metodologia Louvain para identificar comunidades dentro de uma rede ponderada, onde os jogadores eram os nós e os arcos representavam a semelhança numérica entre eles. Aplicado a dados de seis temporadas da NBA (2014–2020), o estudo revelou oito arquétipos de jogadores, como "Supporting Guards" e "Superstar Bigs", superando a classificação tradicional de cinco posições e refletindo a evolução do jogo de um modelo centrado nos postes e no jogo interior para um mais voltado para os bases e o jogo exterior.

Papageorgiou et al. (2024) realizaram um estudo abrangente comparando a eficácia de 14 modelos de machine learning (ML) na previsão do desempenho de jogadores de basquetebol. A análise incluiu 18 estatísticas avançadas e indicadores-chave de desempenho de 90 jogadores de alto desempenho. Cada modelo foi avaliado utilizando o *weighted average percentage error*

(WAPE), desenvolvido a partir do mean absolute percentage error (MAPE) das previsões realizadas. Os resultados indicaram que modelos baseados em árvores, como Extra Trees, Random Forest e Decision Tree, foram os mais eficazes, com Extra Trees a destacar-se ao atingir um WAPE de 34,14%. A abordagem também permitiu melhorar os KPIs selecionados em 3,6% no MAPE para dados não conhecidos, demonstrando o potencial das técnicas de ML para previsões mais precisas e aplicáveis no contexto desportivo.

Todos estes estudos destacam a evolução das técnicas de análise de desempenho no basquetebol, desde abordagens tradicionais baseadas em métricas simples até métodos sofisticados que integram grandes volumes de dados e aprendizagem automática. As investigações de Zhang et al. (2020), Patel (2017), Richardson (2019), Muniz e Flamand (2022) e Papageorgiou et al. (2024) ilustram como metodologias variadas, como clustering, análise de componentes principais e modelos preditivos, proporcionam uma compreensão mais aprofundada dos estilos de jogo, perfis de atletas e fatores que influenciam o desempenho. Além disso, estes trabalhos refletem mudanças nas dinâmicas da NBA, como a crescente complexidade nos papéis dos jogadores e a transição de classificações rígidas para categorias mais flexíveis e multifacetadas, permitindo uma visão mais abrangente das estratégias e do impacto dos jogadores na liga.

4. DATA UNDERSTANDING

A Figura 3 detalha a lista de tarefas associadas à fase de Data Understanding.

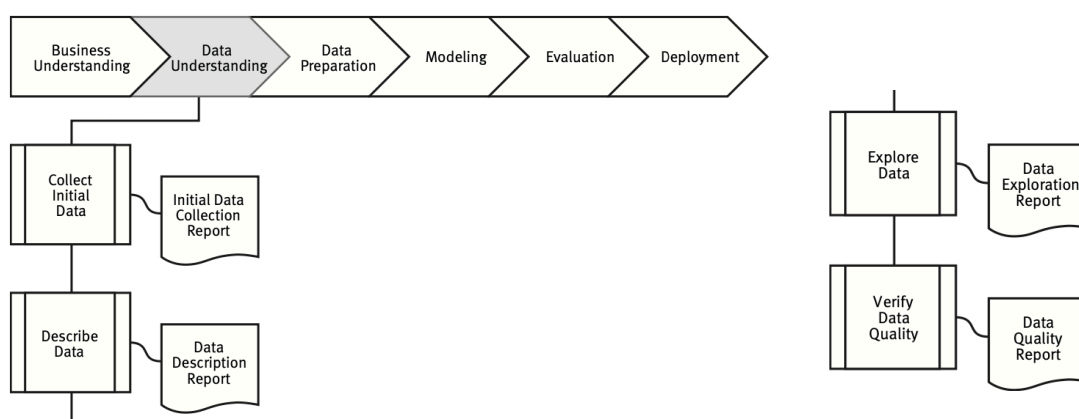


Figura 3 - Descrição das tarefas associadas à fase de Data Understanding

i. Initial Data

Para este projeto, utilizámos dados da temporada regular da NBA 2023/24, recolhidos através de web scraping com recurso à biblioteca do Selenium. As estatísticas diárias dos jogadores, scrapeadas da secção de Daily Leaders do website Basketball Reference¹, foram organizadas em ficheiros CSV mensais, abrangendo dados de outubro de 2023 a março de 2024.

A extração foi realizada acedendo automaticamente aos URLs específicos de cada dia da época, localizando as tabelas com seletores XPath e adicionando a data correspondente a cada registo. Os dados de cada dia foram agregados em DataFrames mensais e exportados para CSVs.

No final, com recurso às bibliotecas glob e pandas foram agregados todos os dados num único DataFrame, resultando num conjunto unificado e organizado com as estatísticas recolhidas da performance individual de cada jogador durante a época regular de NBA 2023/24. Cada linha do DataFrame gerado representa as estatísticas obtidas por esse jogador nesse determinado jogo.

¹ <https://www.basketball-reference.com/friv/dailyleaders.fcgi?month=11&day=18&year=2010&type=all>

Estes dados iniciais representam uma base essencial para as análises realizadas neste trabalho, proporcionando uma visão abrangente sobre as estatísticas dos jogadores ao longo da temporada regular. Estes registos incluem diversas informações relacionadas com o jogador, a equipa e o jogo, podendo ser demonstradas e explicadas tal como se encontra na Tabela 1.

Variável	Explicação da variável
Player	Nome do jogador que participou no jogo
Tm	Indica a equipa à qual o jogador pertence
vs	Indica se a equipa do jogador jogou em casa ("Nan") ou jogou fora ("@")
Opp	Indica a equipa adversária que o jogador enfrentou
W/L	Indica se a equipa do jogador venceu ("W") ou perdeu ("L") o jogo
MP	Minutos jogados pelo jogador no jogo
FG	Número de cestos de campo convertidos pelo jogador
FGA	Número de tentativas de cestos de campo feitas pelo jogador
FG%	Percentagem de eficácia nos cestos de campo
3P	Número de cestos de três pontos convertidos pelo jogador
3PA	Número de tentativas de cestos de três pontos feitas pelo jogador
3P%	Percentagem de eficácia nos cestos de três pontos
FT	Número de lançamentos livres convertidos pelo jogador
FTA	Número de tentativas de lançamentos livres pelo jogador
FT%	Percentagem de eficácia nos lançamentos livres
ORB	Número de ressaltos ofensivos conquistados (na área do adversário)
DRB	Número de ressaltos defensivos conquistados (na sua própria área)
TRB	Número total de ressaltos conquistados
AST	Número de assistências realizadas
STL	Número de roubos de bola efetuados
BLK	Número de desarmes de lançamento efetuados
TOV	Número de perdas de bola cometidas pelo jogador
PF	Número de faltas cometidas (Personal Fouls)
PTS	Pontos totais marcados pelo jogador no jogo
+/-	Diferença no marcador enquanto o jogador esteve em campo
GmSc	Métrica numérica que resume o desempenho global do jogador no jogo, considerando várias estatísticas
Date	Data em que o jogo ocorreu

Tabela 1 - Dicionário de variáveis

Para facilitar a análise exploratória de dados, procedeu-se a algumas tarefas de processamento na fase de Data Understanding, nomeadamente no que diz respeito à criação de uma tabela resumo com os dados médios de cada jogador para cada uma das variáveis numéricas disponíveis. As secções seguintes da fase de Data Understanding relativas à Data Quality e Data Exploration dizem respeito precisamente a estes dados médios da temporada regular já processados.

ii. Data Quality

Na análise de dados, outliers são frequentemente identificados como observações que se distanciam significativamente do comportamento geral dos dados. Contudo, é essencial considerar o contexto e a natureza do dataset antes de se decidir remover ou modificar estes valores.

No caso em análise, os boxplots gerados, que podem ser vistos na Fig. 4, evidenciam a presença de valores que poderiam, à primeira vista, ser considerados outliers. No entanto, ao investigar a origem dessas observações, constatamos que estas correspondem a desempenhos de jogadores de elite, como LeBron James, entre outros atletas de alta performance. Esses valores não são erros ou anomalias, mas sim reflexos de desempenhos reais que enriquecem o dataset e são fundamentais para o estudo em questão.

Assim, optou-se por manter estas observações no dataset, uma vez que removê-las ou ajustá-las resultaria numa distorção dos dados e comprometeria a validade das análises e conclusões. Além disso, a presença de outliers contribui para a identificação de padrões únicos e estratégias específicas no desempenho dos jogadores.

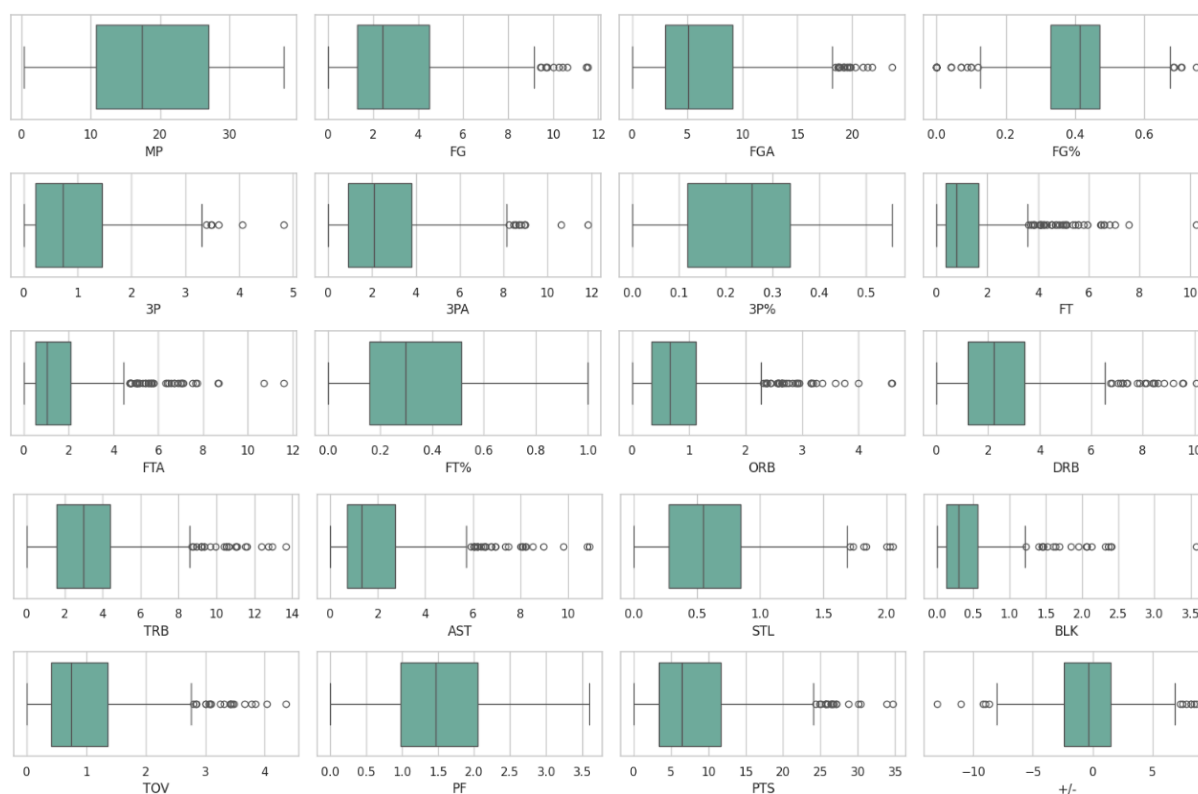


Figura 4 - Boxplots de todas as variáveis numéricas

iii. Data Exploration

Na fase inicial da exploração de dados, é fundamental compreender as distribuições das principais variáveis e estas podem ser vistas na Figura 5. A análise das distribuições revelou alguns padrões interessantes nas estatísticas dos jogadores. A variável *MP* (*Minutos Jogados*) apresenta uma distribuição relativamente uniforme, com maior concentração de jogadores no intervalo de 20 a 30 minutos por jogo, refletindo uma gestão equilibrada do tempo em campo durante a época regular.

Nas variáveis *FG* (*Cestos de Campo Convertidos*) e *PTS* (*Pontos*), observa-se uma forte relação, ambas apresentando distribuições enviesadas à direita. A maioria dos jogadores marca poucos pontos e converte poucos cestos, mas as caudas direitas destacam atletas com contribuições ofensivas significativas. Já a *FG%* (*Porcentagem de Conversão de Cestos de Campo*) apresenta uma distribuição normal em torno de 50%, o que é representativo da eficácia média dos jogadores na liga.

Por outro lado, a análise das variáveis *ORB* (*Ressaltos Ofensivos*) e *DRB* (*Ressaltos Defensivos*) mostra uma discrepância clara: os ressaltos defensivos são muito mais frequentes do que os ofensivos. Este fenómeno pode ser explicado pela disposição tática dos jogadores ofensivos, que muitas vezes utilizam um esquema “five out”, mantendo-se distantes do cesto. Quando ocorre um lançamento, os jogadores ofensivos ficam geralmente mais longe da bola em caso de ressalto, já que os adversários estão posicionados entre o cesto e o jogador ofensivo. Além disso, os jogadores defensivos utilizam a técnica de “box out”, bloqueando fisicamente os adversários para garantir o ressalto defensivo, limitando as oportunidades de ressaltos ofensivos.

A variável *TOV* (*Perdas de Bola*) sugere que embora a maioria dos jogadores cometa poucas perdas de bola, aqueles mais ativos na criação de jogadas tendem a errar mais devido à maior exposição. Por fim, a variável *+/-* (*Impacto no Marcador*) apresenta uma distribuição simétrica e centrada em torno de zero, mostrando assim um equilíbrio geral entre jogadores com impacto positivo e negativo no marcador.

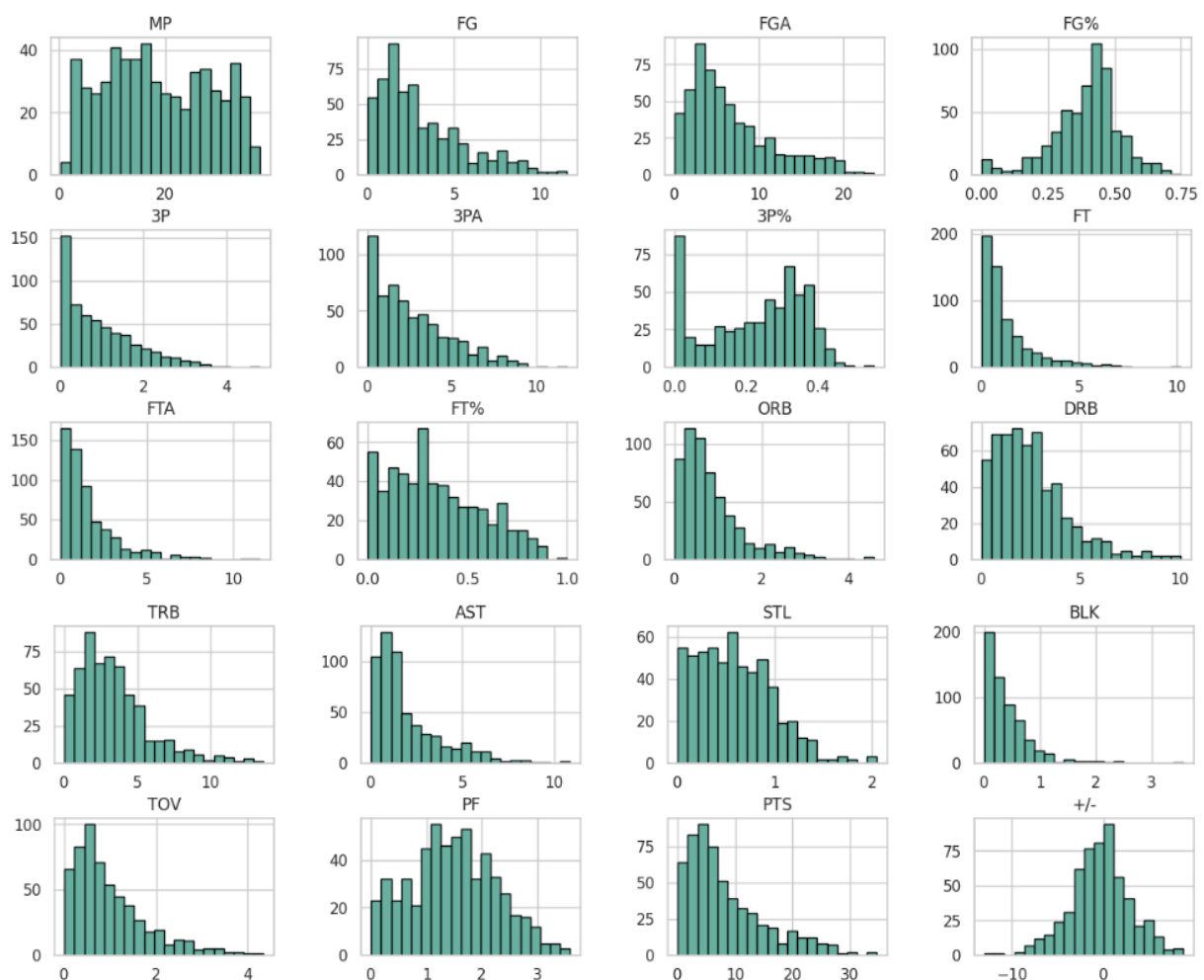


Figura 5 - Distribuições das variáveis numéricas

5. DATA PREPARATION

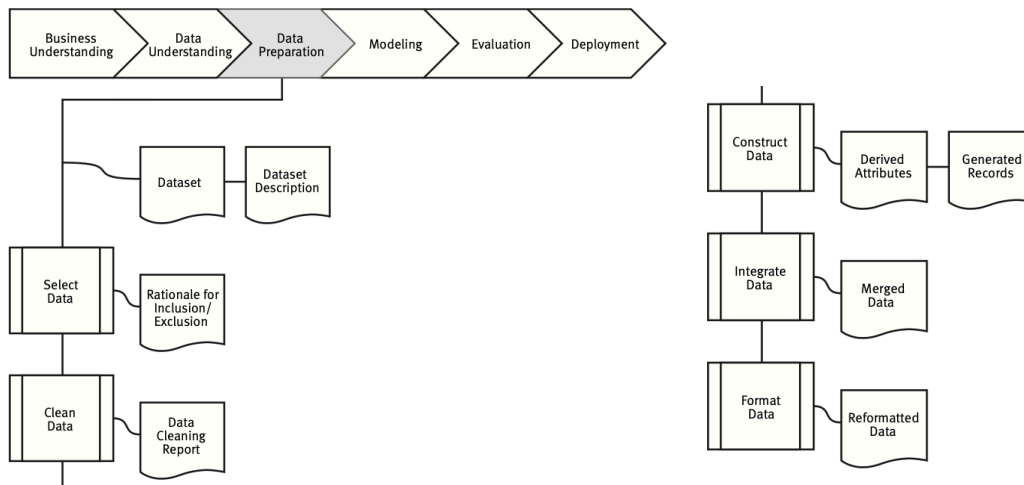


Figura 7 - Descrição das tarefas associadas à fase de Data Preparation

i. Valores omissos

Durante a fase de preparação de dados, foi realizado o tratamento dos valores omissos. Nas percentagens de lançamento, as colunas relacionadas com a eficácia de arremessos (FG%, 3P% e FT%) apresentavam valores nulos para jogadores que não realizaram tentativas. Para corrigir este problema, esses valores foram imputados com uma constante 0, que reflete corretamente a ausência de tentativas de lançamento. Na estatística de +/-, que indica a diferença de pontuação da equipa enquanto o jogador está em campo, também se verificaram valores nulos. Para os jogadores que não jogaram (MP = 0), imputou-se a constante 0, enquanto para os restantes foi utilizada a mediana do +/- dos jogos em que participaram, dado o número reduzido de observações com valores omissos para esta variável.

ii. Feature engineering e transformação de variáveis

A fase de feature engineering foi conduzida com o objetivo de enriquecer os dados, criando variáveis relevantes que ajudassem a identificar padrões no desempenho dos jogadores. Estas variáveis foram pensadas para facilitar não só a análise exploratória, mas também os passos subsequentes de modelação, que inclui a análise de componentes principais e a formação de clusters através do k-means, além da interpretação de resultados para perfilar os diferentes grupos/clusters.

Um dos primeiros passos foi transformar a variável MP (minutos jogados), originalmente registada no formato MM:SS, para um formato decimal, de forma a garantir consistência nas análises numéricas e permitir a inclusão desta métrica de forma mais eficaz na identificação de padrões e grupos de jogadores com base no tempo de jogo.

Para capturar as diferenças no desempenho dos jogadores com base no local dos jogos, foi criada uma variável binária que distingue entre jogos em casa e fora. A partir dela, foram geradas métricas como:

- *Home_Games_Count*: Número de jogos em casa, que é útil para identificar jogadores que têm maior participação, melhor performance ou maior destaque em jogos realizados no seu ambiente habitual ou com o apoio da bancada.
- *Away_Games_Count*: Número de jogos fora, que ajuda a perceber consistência ou variações de desempenho fora do ambiente habitual de cada jogador.

Estas métricas fornecem insights importantes sobre o contexto de jogo e ajudam a distinguir jogadores que podem apresentar diferenças significativas em jogos dentro e fora de casa.

Para aprofundar a análise dos jogos em casa, foram calculadas métricas como:

- *Home_Wins_Count*: Número de vitórias em casa, indicando a contribuição do jogador em cenários que se traduziram na vitória da sua equipa.
- *Home_Losses_Count*: Diferença entre o número total de jogos em casa e as vitórias, representando as derrotas e permitindo entender melhor a influência do jogador em resultados negativos da sua equipa.

Outro conjunto de métricas importantes que se optou por criar teve que ver com a identificação dos adversários contra os quais cada jogador teve desempenhos de destaque ou passíveis de ser assinalados:

- *Top_Opponent*: Adversário contra o qual o jogador marcou o maior número de pontos.
- *Top_Rebound_Opponent*: Adversário contra o qual obteve mais ressaltos.
- *Top_Assist_Opponent*: Adversário contra o qual realizou mais assistências.

Estas variáveis ajudam a captar a capacidade de um jogador e a forma como este se destaca contra adversários específicos e capta as nuances do seu desenvolvimento contra os planos táticos e técnicos de certas equipas/adversários. Juntamente com isto, foram criadas as métricas numéricas correspondentes à soma de pontos do jogador contra o seu principal oponente (*Points_vs_Top_Opponent*), à soma de ressaltos do jogador contra o adversário em que teve melhor desempenho (*Rebounds_vs_Top_Opponent*), e à soma de assistências (*Assists_vs_Top_Opponent*).

Estas informações, tal como já foi referido, são particularmente importantes para as componentes principais e para o clustering, pois permitem identificar grupos de jogadores com características semelhantes, como consistência em jogos contra adversários fortes e tornam-se essenciais na perfilagem dos clusters, ao descrever os fatores que diferenciam os grupos.

6. MODELING

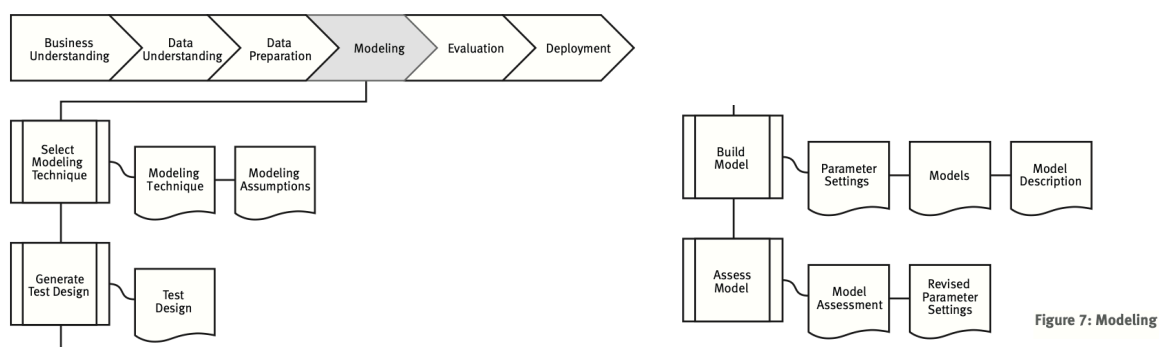


Figura 8 - Descrição das tarefas associadas à fase de Modeling

Na fase de modelação, utilizou-se a Análise de Componentes Principais para reduzir a dimensionalidade do conjunto de dados e identificar os componentes que explicam a maior parte da variância nos dados. Posteriormente, aplicou-se o *k-means* para agrupar jogadores com características semelhantes com base nos componentes principais selecionados.

i. Análise de componentes principais (PCA)

Antes da aplicação da PCA, as variáveis numéricas foram padronizadas com recurso ao StandardScaler da biblioteca Sci-kit learn, garantindo que todas as variáveis contribuíssem igualmente para o cálculo dos componentes principais, independentemente das suas escalas originais. Note-se que no dataset estão presentes % de arremesso que variam entre 0 e 1 e total de pontos marcados, a título de exemplo, que supera largamente a escala das % de arremesso. Nesse sentido, esta etapa é essencial para evitar que variáveis com magnitudes muito maiores dominem a análise.

Inicialmente, a PCA foi aplicada com o número de componentes igual ao número total de variáveis numéricas, permitindo uma análise completa da variância explicada por cada componente. Os gráficos presentes na Fig. 9, foram utilizados para identificar o número ideal de componentes com base na variância explicada. Observou-se que cinco componentes principais explicam aproximadamente 86,8% da variância total, uma proporção que se considerou suficiente para capturar a maior parte da informação relevante nos dados.

Além disso, o critério de Kaiser, que considera apenas as componentes com eigenvalues superiores a 1, foi também considerado e tal resultou em 4 componentes principais, conforme se pode ver na Tabela 2. No entanto, optou-se por incluir 5 componentes para garantir uma explicação mais robusta da variância.

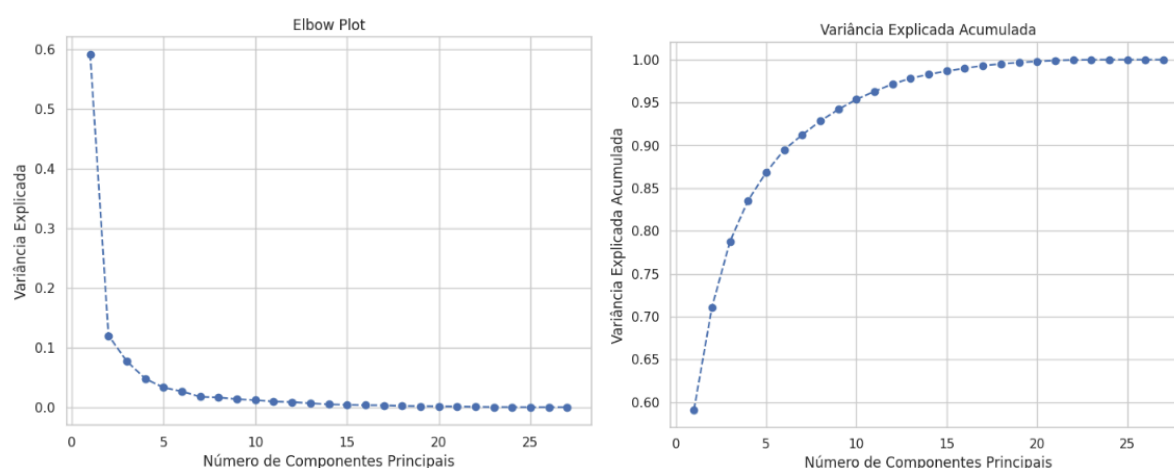


Figura 9 - Variância explicada dos componentes principais

Componente	Eigenvalue	Kaiser Criterion (> 1)
1	15,9906	✓
2	3,2425	✓
3	2,0791	✓
4	1,2916	✓
5	0,8859	✗
...		
27	0,0000	✗

Tabela 2 - Aplicação do Critério de Kaiser para seleção de número de componentes

As cargas (ou *loadings*) das cinco componentes principais foram analisadas para interpretar quais as variáveis originais que contribuíram mais significativamente para cada componente. A Tabela 3 resume as cargas das variáveis em cada componente principal.

	PC1	PC2	PC3	PC4	PC5
MP	0.2375	-0.024941	0.008273	-0.090479	0.152159
FG	0.237611	-0.04097	-0.139871	0.014682	0.007537
FGA	0.23334	-0.116052	-0.14299	-0.035092	0.033698
FG%	0.168257	0.214695	0.135566	-0.002284	0.091232
3P	0.175349	-0.328508	0.060247	-0.027852	0.323105
3PA	0.175259	-0.330925	0.026042	-0.068662	0.301402
3P%	0.146396	-0.290194	0.189745	-0.019814	0.386205
FT	0.209942	-0.034696	-0.268919	0.077303	-0.192777
FTA	0.210979	0.009648	-0.268575	0.072136	-0.204015
FT%	0.217862	-0.01827	-0.136143	-0.031986	-0.094449
ORB	0.11808	0.447434	-0.040073	-0.014182	0.087718
DRB	0.207079	0.241449	-0.074032	0.014532	0.103564
TRB	0.192469	0.319616	-0.068151	0.006601	0.105067
AST	0.195956	-0.159018	-0.15475	0.033818	-0.248963
STL	0.186745	-0.042176	-0.039122	0.014431	0.211172
BLK	0.126478	0.331546	-0.022402	0.032406	0.27701
TOV	0.218903	-0.04733	-0.207085	-0.082715	-0.12594
PF	0.191576	0.149577	0.045443	-0.175316	0.212487
PTS	0.237628	-0.079206	-0.148395	0.022802	0.007899
+/-	0.066269	-0.009628	0.112158	0.794815	0.049952
Home_Games_Count	0.188051	0.002386	0.428726	-0.048675	-0.197142
Away_Games_Count	0.188296	0.01398	0.410838	-0.034741	-0.189484
Home_Wins_Count	0.161579	-0.002636	0.379689	0.346193	-0.089763
Home_Losses_Count	0.154822	0.006476	0.342076	-0.413653	-0.238971
Points_vs_Top_Opponent	0.23527	-0.090999	-0.023015	0.043779	-0.09054
Rebounds_vs_Top_Opponent	0.19813	0.277276	0.080404	0.008805	-0.031342
Assists_vs_Top_Opponent	0.203247	-0.144344	-0.014802	0.055322	-0.331329

Tabela 3 - Loadings das componentes principais

As componentes principais representam diferentes dimensões do desempenho dos jogadores com base nas variáveis analisadas. Cada componente foi interpretada de acordo com as variáveis que mais contribuíram para ela.

A **PC1** está fortemente relacionada à **produção ofensiva de um jogador**. As variáveis como minutos jogados (MP), arremessos convertidos (FG), tentativas de arremesso (FGA), três pontos convertidos (3P) e pontos (PTS) sugerem que esta componente captura tanto a eficiência quanto o volume de jogo ofensivo.

A **PC2** é fortemente correlacionada com o **desempenho defensivo e os ressaltos**. Variáveis como ressaltos ofensivos (ORB), defensivos (DRB), totais (TRB) e bloqueios (BLK) indicam que esta componente reflete a capacidade de defesa e controlo do garrafão.

A **PC3** está associada ao **desempenho em jogos disputados em casa vs fora**. As variáveis como o número de jogos em casa (Home_Games_Count), jogos fora (Away_Games_Count), vitórias em casa (Home_Wins_Count) e derrotas em casa (Home_Losses_Count) capturam o impacto das condições de jogo em casa no desempenho dos jogadores.

A **PC4** mede o **impacto geral de um jogador no jogo**, sendo fortemente influenciada pela diferença de pontos enquanto o jogador está em campo (+/-) e pelas vitórias e derrotas em casa. Este componente reflete a contribuição global de um jogador para o sucesso da equipa, sendo interpretada como o impacto geral do jogador no jogo.

Por fim, a **PC5** combina as variáveis relacionadas com a **eficiência nos arremessos** de três pontos (3P%), na linha de lance livre (FT%) e assistências contra adversários fortes (Assists_vs_Top_Opponent), representando assim a eficiência no ataque e a habilidade técnica e tática para criar jogadas.

ii. k-Means Clustering

O k-Means foi aplicado para identificar agrupamentos/clusters significativos entre os jogadores com base nas cinco componentes principais extraídas anteriormente. Primeiramente, as componentes principais (PC1, PC2, PC3, PC4 e PC5) foram adicionadas ao DataFrame original, servindo como base para o agrupamento. Em seguida, foram testados diferentes valores de k, variando de 2 a 10, para determinar o número ideal de clusters. Para isso, foi utilizado o Silhouette Score como métrica de qualidade, e os resultados podem ver-se nos gráficos da Fig. 10 que mostram a distribuição dos clusters em função das duas primeiras componentes principais (PC1 e PC2). Note-se que as duas primeiras componentes representam cerca de 71% da variabilidade dos dados. O Silhouette Score foi incluído nos gráficos para facilitar a análise da qualidade do agrupamento.

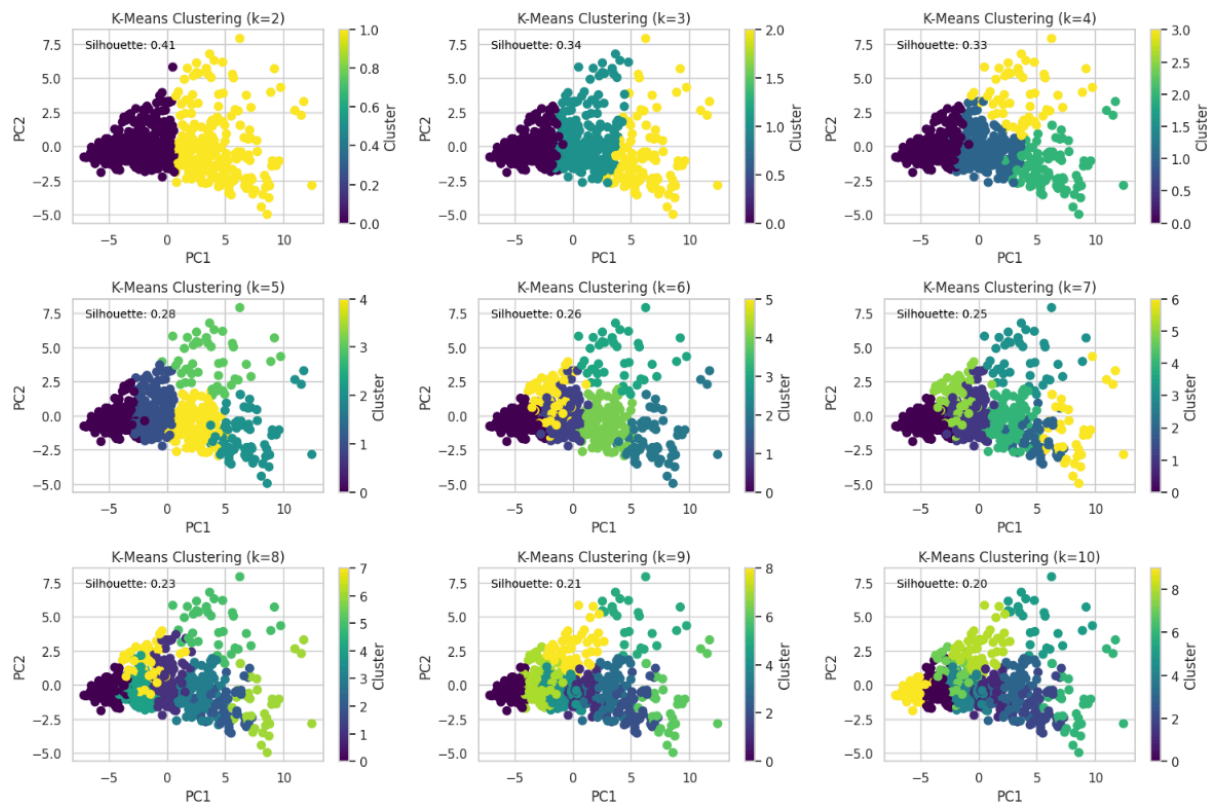


Figura 10 - Visualização gráfica de diferentes clusters formados com k-Means para diferentes valores de K

Além disso, foi aplicado o método do cotovelo (Elbow Method) para identificar o valor de k mais adequado. O gráfico gerado pelo método, apresentado na Fig. 11, revelou que o valor ideal de k é 5, baseado no ponto de inflexão do gráfico da soma dos quadrados das distâncias dentro dos clusters (WCSS). Após essa análise, o k -Means foi executado com $k=5$, e o Silhouette Score para esse valor de k foi calculado, resultando em um valor de 0,28. Embora esse valor seja razoável, a melhor pontuação de Silhouette foi observada para $k=2$ (0,41), indicando uma separação um pouco melhor entre os clusters. No entanto, optou-se por manter $k=5$, já que se considerou que o número de clusters gerado se apresentava como mais adequado para a análise.

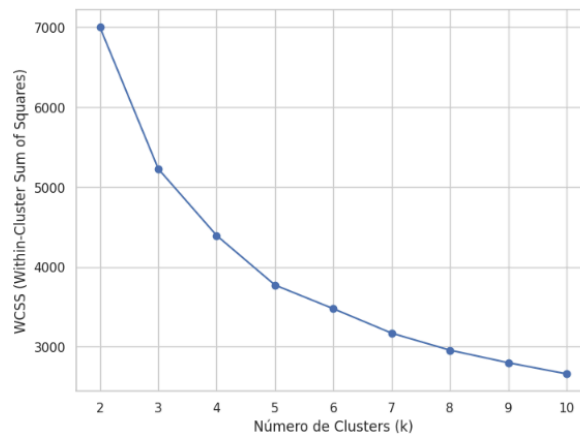


Fig. 11 - Método do cotovelo para seleção do número de clusters (k)

7. EVALUATION

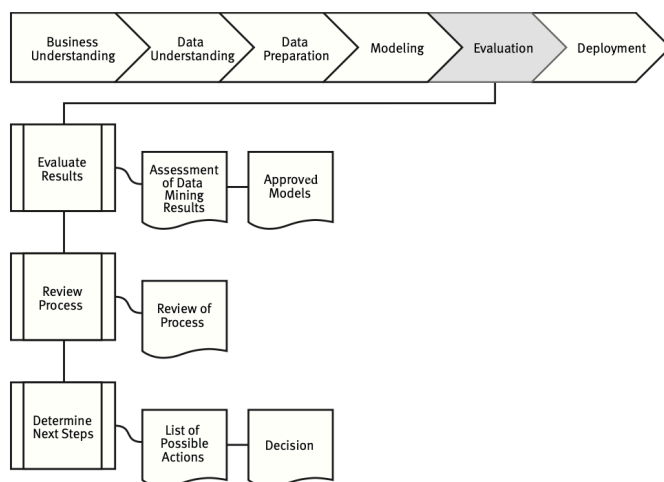


Figura 12 - Descrição das tarefas associadas à fase de Evaluation

Na fase de Evaluation do CRISP-DM, o objetivo é analisar os resultados dos clusters gerados e validar a sua utilidade. Neste caso, a análise de clusters tem como objetivo identificar padrões de desempenho entre jogadores da NBA, classificando-os com base nas suas estatísticas e características de jogo. A avaliação dos clusters ajudará a entender melhor os diferentes tipos de jogadores e pode fornecer insights sobre como as equipas podem otimizar as suas estratégias de composição e utilização dos atletas, bem como permite entender as características e nuances que definem o basquetebol dos Estados Unidos. Na Fig. 13 cada ponto representa um jogador, colorido de acordo com o cluster ao qual foi atribuído e estão identificados os principais clusters que caracterizam a NBA.

A observação de que há uma maior proximidade entre jogadores estrela e jogadores ofensivos na NBA pode ser contextualizada com uma citação de Luka Dončić, que afirmou que "é mais fácil marcar pontos na NBA do que na Europa" (Dončić, 2023). Isso pode ser explicado pelo estilo de jogo mais orientado para o ataque característico da NBA, onde o talento ofensivo é mais valorizado. Dončić destaca que a defesa na NBA é frequentemente menos organizada de forma tática e mais focada em impedir jogadas individuais, enquanto que na EuroLeague é muito mais estruturada e rígida, tornando mais difícil para os jogadores ofensivos dominarem de maneira tão clara. Este contraste revela como o jogo da NBA, com suas permissões para mais movimentação livre e ênfase no ataque, favorece um estilo de jogo mais fluido e dinâmico, o que cria um ambiente onde jogadores ofensivos se destacam mais. Assim, a proximidade entre jogadores estrela e ofensivos nas análises de clusters pode refletir as diferenças fundamentais entre essas duas ligas, reforçando a ideia de que a NBA é um palco onde as habilidades ofensivas são mais exaltadas.

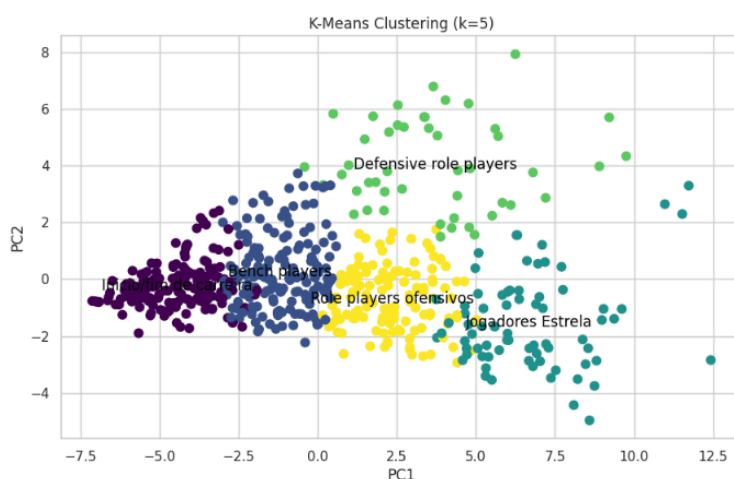


Fig. 13 - Visualização gráfica dos clusters gerados com k-Means (com k=5)

A caracterização destes clusters teve por base a análise dos gráficos presentes na Figura 14 e que permitem perfilar os clusters através da seguinte caracterização cuja análise se pode sumarizar abaixo:

- **Cluster 0: "Jogadores Menos Ativos / Início ou Fim de Carreira"**
 - a) Atividade: Baixa média de minutos por jogo (7.75), com estatísticas abaixo da média.
 - b) Ofensiva: Menor envolvimento ofensivo, com poucos pontos, assistências e lançamentos.
 - c) Defesa: Contribuição limitada em ressaltos e bloqueios.
- **Cluster 1: "Jogadores de Papel Secundário/Bench Players"**
 - a) Atividade: Média de 15.69 minutos por jogo.
 - b) Ofensiva: Baixo volume ofensivo. Apenas 5.41 pontos de média.
 - c) Defesa: Contribuem razoavelmente em ressaltos (com mais destaque para ressaltos defensivos) e assistências.
- **Cluster 2: "Jogadores Estrela"**
 - a) Atividade: Muito ativos com média de 33.56 minutos por jogo.
 - b) Ofensiva: Grande destaque em pontos (21.98), assistências e boa percentagem de lançamento (FG% de 46.8%).
 - c) Defesa: Forte em ressaltos, com bom desempenho nas duas extremidades do jogo.
- **Cluster 3: "Defensive Role Players"**
 - a) Atividade: Média de 26.91 minutos, com bom desempenho equilibrado.
 - b) Ofensiva: Boa produção de pontos e ressaltos ofensivos.

c) Defesa: Boa contribuição defensiva, especialmente em ressaltos.

- **Cluster 4: "Role Players Ofensivos"**

- a) Atividade: Média de 25.75 minutos, com bom desempenho ofensivo e defensivo.
- b) Ofensiva: Boa produção em pontos (11.28) e lançamentos.
- c) Defesa: Contribuição sólida em ressaltos e assistências.

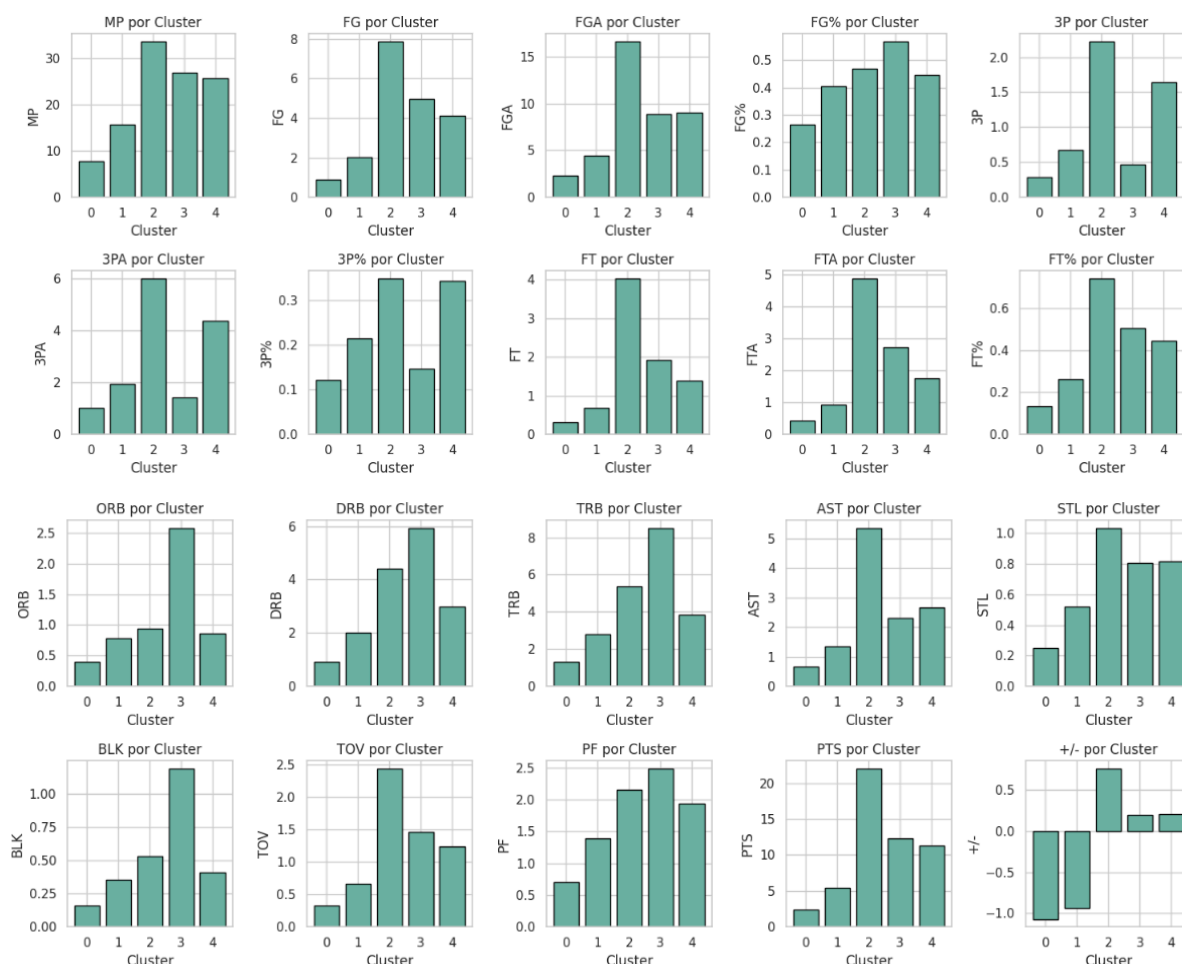


Figura 14 - Caracterização das estatísticas médias em cada cluster

A Figura 15 destaca as equipas com o maior número de jogadores em cada cluster. Assim, é possível identificar tendências específicas dentro de cada grupo, como estilos de jogo predominantes ou características comuns entre os jogadores de uma mesma equipa.

- **Cluster 0 (Jogadores Menos Ativos / Início ou Fim de Carreira):**

Equipas como **Raptors**, **Pelicans** e **76ers** têm jogadores com menor envolvimento em minutos e com estatísticas abaixo da média, provavelmente com veteranos ou jogadores em fase de transição de carreira. Essas equipas têm uma contribuição limitada ofensiva e defensivamente. Isso reflete-se na tabela classificativa onde apenas os 76ers conseguiram chegar a lugares de playoff mas a perderem logo na primeira fase.

- **Cluster 1 (Jogadores de Papel Secundário/Bench Players):**

Wizards, **Grizzlies** e **Spurs** apresentam uma média de minutos por jogo mais baixa, com uma contribuição defensiva razoável, mas com um volume ofensivo muito limitado. As

equipas que aparecem neste cluster fazem sentido pois nesta época em questão as equipas decidiram entrar em fase de *Rebuild*, isto é, assumiram uma postura menos vitoriosa para ficarem com *picks do draft* melhores², o que resulta em rodar mais o plantel de maneira também a entenderem quem tem potencial para se manter na equipa. Os Wizards e Spurs assumem esta postura por falta de qualidade no plantel, já os Grizzlies viram-se obrigados a rodar plantel mais devido à grande quantidade de jogadores lesionados.

- **Cluster 2 (Jogadores Estrela):**

Celtics, Knicks e **Portland** têm estrelas no seu elenco, com jogadores a fazerem grandes médias de pontos, assistências e ressaltos, com uma forte capacidade de impactar ambos os lados do jogo. Estas equipas têm uma base sólida com jogadores de topo que são a espinha dorsal do sucesso da equipa. Celtics e Knicks tiveram recordes muito positivos, já Portland não se viu ser tão sucedido, o que pode indicar que os jogadores estrelas estão lá, falta apenas rodeá-los com jogadores melhores nos outros papeis de jogo.

- **Cluster 3 (Defensive Role Players):**

Rockets, Hawks e **Nets** têm jogadores que contribuem de forma equilibrada, com um bom desempenho defensivo, mas com uma capacidade ofensiva menos destacada. São jogadores que ajudam a manter a equipa competitiva em aspectos defensivos. Estas equipas mantiveram-se a meio da tabela, o que nos indica que apenas jogadores mais defensivos não chegam para patamares mais elevados.

- **Cluster 4 (Role Players Ofensivos):**

Pacers, Grizzlies e **Celtics** têm jogadores que são eficazes em ambos os lados do campo, com destaque para a produção ofensiva, mas também com boa contribuição defensiva. Estes jogadores são fundamentais no equilíbrio da equipa, com bons desempenhos tanto em pontos quanto em ressaltos e assistências. Este cluster já tem uma dinâmica diferente dos anteriores, uma vez que as três equipas aqui inseridas se encontraram em cenários diferentes umas das outras. Os Celtics, como já dito, tiveram um recorde muito positivo, os Pacers ficaram a meio da tabela e os Grizzlies na parte mais baixa da mesma. Isto indica-nos que qualquer equipa pode ter jogadores com perfil mais atacante, mas tê-los não faz com que haja garantias de sucesso.

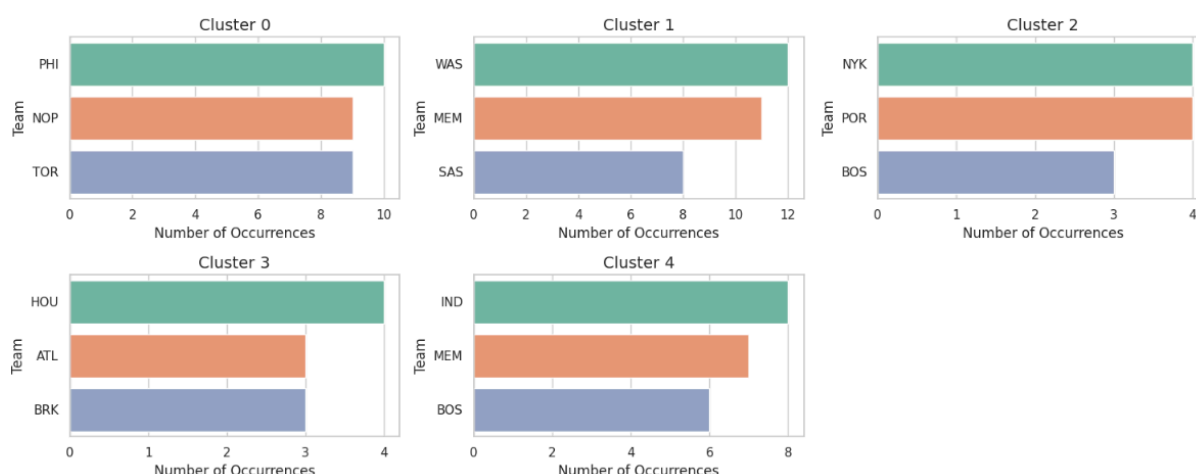


Figura 15 - Equipas com o maior número de jogadores em cada cluster

Para analisar a performance global de cada jogador, foi calculado um score ponderado, levando em consideração diversas estatísticas individuais, com diferentes pesos para refletir a importância relativa de cada uma delas. A fórmula utilizada para calcular o score foi a seguinte:

² Na NBA os piores classificados por época recebem as melhores picks do draft de forma geral.

$$df['Score'] = (2 * df['PTS'] + 2 * df['AST'] + 0.5 * df['TRB'] + 2 * df['BLK'] + 2 * df['STL'] - 2 * df['TOV']) / df['MP']$$

Neste cálculo, os PTS (Pontos) e AST (Assistências) receberam o peso de 2, pois são as métricas principais para avaliar o desempenho ofensivo e a criação de jogadas. Os TRB (Ressaltos) tiveram um peso de 0.5, dado que, apesar de importantes, têm uma relevância menor em comparação com pontos e assistências. As métricas BLK (Bloqueios) e STL (Roubos de bola) receberam o peso de 2, uma vez que são indicadores significativos de desempenho defensivo. As TOV (Perdas de bola) foram subtraídas, com um peso negativo de 2, uma vez que prejudicam o desempenho do jogador. Por fim, o score foi normalizado pelos MP (Minutos jogados), garantindo que o desempenho fosse avaliado proporcionalmente ao tempo em campo.

A partir desses scores, foi possível identificar o Top 10 jogadores por cluster, permitindo uma análise detalhada dos jogadores que se destacam dentro de cada grupo, com base na sua performance global calculada pela fórmula mencionada. Na Figura 16 são apresentados 5 gráficos, cada um correspondendo a um cluster em que cada painel compara o score ao longo da época regular dos jogadores dentro desse cluster. Analisando os resultados apresentados nestes gráficos, podemos obter padrões interessantes que nos permitem compreender melhor a contribuição de cada grupo de atletas.

No cluster 0 temos os jogadores menos ativos, que se encontram em início ou fim de carreira. Nomes como Markquis Nowell e Isaiah Brockington encaixam-se neste perfil, sugerindo que poderão estar a iniciar ou a terminar as suas carreiras profissionais, mesmo tendo um score bastante positivo, que neste caso podem ter sido consequência de poucos jogos efetuados e boas estatísticas em poucos minutos jogados, isto porque normalmente estes jogadores entram quando jogos já não são importantes ou em jogos em que o resultado tem muita diferença entre as duas equipas.

Já no cluster 1, temos os jogadores de papel secundário/*bench players*. Nomes como Isaiah Jackson, Brandon Clarke e Neemias Queta integram este grupo. Tratam-se de jogadores que são muito importantes para o banco das suas equipas, e que apresentam scores muito semelhantes entre si. Importa reforçar que este é o grupo onde há maior similaridade entre o seu top de atletas.

Por outro lado, no cluster 2 encontramos os jogadores estrela. Este é o grupo dos principais destaques da liga como Joel Embiid, Luka Dončić e Giannis Antetokounmpo. De notar que a partir da classificação com a fórmula de Score dos jogadores ao longo da época regular, tirando da análise a quantidade de jogos efetuados e os intangíveis de um jogador que não se medem a partir de estatísticas, podemos dizer que o premiado³ com Most Valuable Player (MVP) da Liga 2023/24, Nikola Jokic, aparece apenas em 5º lugar, sendo que o melhor, a partir do Score, Joel Embiid, ficou de fora dos selecionados.

O cluster 3 reúne os melhores jogadores defensivos. Nomes como Victor Wembanyama, Anthony Davis e Domantas Sabonis representam bem este perfil e são top 3 a nível de Score. Importa reforçar que o jogador que recebeu o prémio de Defensive Player of the Year 2023/24, Rudy Gobert, não aparece sequer nos primeiros 10 em termos de Score, o que pode ser explicado pelo fraco desempenho ofensivo do jogador ou por não se considerarem elementos intangíveis nesta classificação. A par e passo, importa reforçar que Victor Wembanyama, jogador com o Score mais elevado neste cluster, recebeu o prémio de Rookie of the Year e terminou em 2º lugar para Defensive Player of the Year, atrás de Gobert.

Finalmente, no cluster 4 temos os role players ofensivos. Estes atletas demonstram um bom desempenho ofensivo e jogadores como T.J. McConnell, Malik Monk e Cam Whitmore encaixam-se

³ A atribuição de prémios 2023/24, os jogadores selecionados e as respectivas pontuações podem ser consultadas em: https://www.basketball-reference.com/awards/awards_2024.html

nesta categoria. É importante referir que Malik Monk, o número dois na classificação no gráfico a partir do Score ao longo da temporada regular, esteve entre os 3 nomeados para Kia NBA Sixth Man of the Year, perdendo para Naz Reid, jogador dos Minnesota Timberwolves.

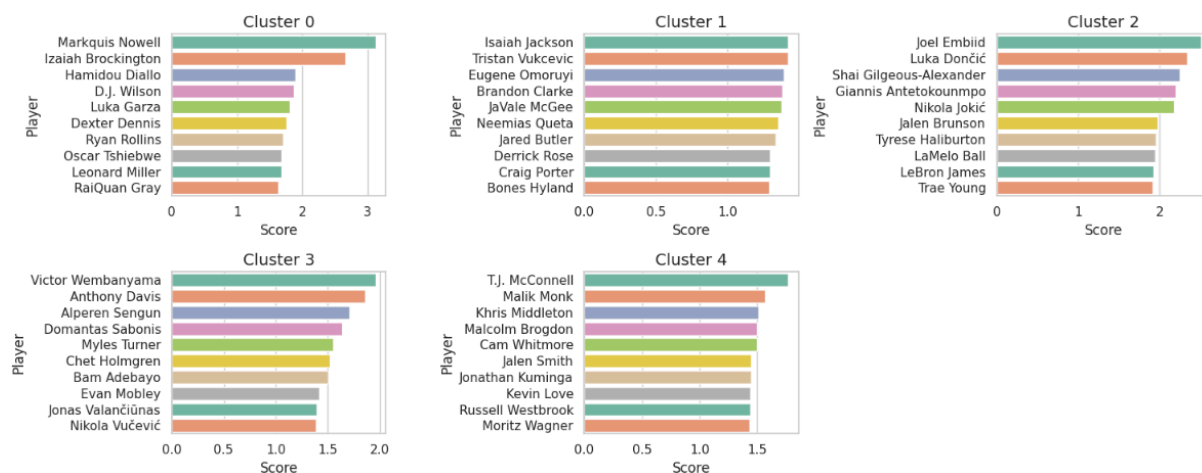


Figura 16 - Melhores jogadores presentes em cada cluster através da fórmula de Score calculada

Como discutido na análise do Cluster 2, observa-se que, apesar de certos jogadores serem considerados os melhores da liga, a classificação pelo Score ao longo da temporada nem sempre reflete a atribuição dos prémios. Na Figura 17, Nikola Jokić, premiado com o MVP, apresenta uma linha de Score discreta, com alguns picos positivos em comparação com outros jogadores. Isto indica que o Score calculado com base apenas em estatísticas quantitativas nem sempre reflete com precisão a contribuição real do jogador para o sucesso da sua equipa. Analistas, como Chris Herring, NBA insider, abordam este ponto num artigo publicado pela ESPN, destacando a influência de Jokić na performance da sua equipa para a época atual 2024/25 (Herring, 2024).

A discrepância entre o ranking baseado no Score e a atribuição do prémio real de MVP mostra que a atribuição dos prémios leva em consideração não apenas os números, mas também a influência global do jogador na dinâmica da equipa e no sucesso coletivo e outros fatores intangíveis como a liderança, o impacto estratégico e a influência sobre os companheiros de equipa.

O gráfico da figura 17 mostra que Joel Embiid, até final de janeiro, era um forte candidato a MVP, mas ao sofrer uma lesão grave acabou por ficar de fora da corrida ao prémio. Por fim, pode concluir-se, com base nas métricas mensais, que Luka Dončić foi o jogador com o desempenho mais constante e de alto nível ao longo da temporada regular através da métrica de Score e este acabou na 3ª posição na classificação de MVP, atrás de Shai Gilgeous-Alexander.

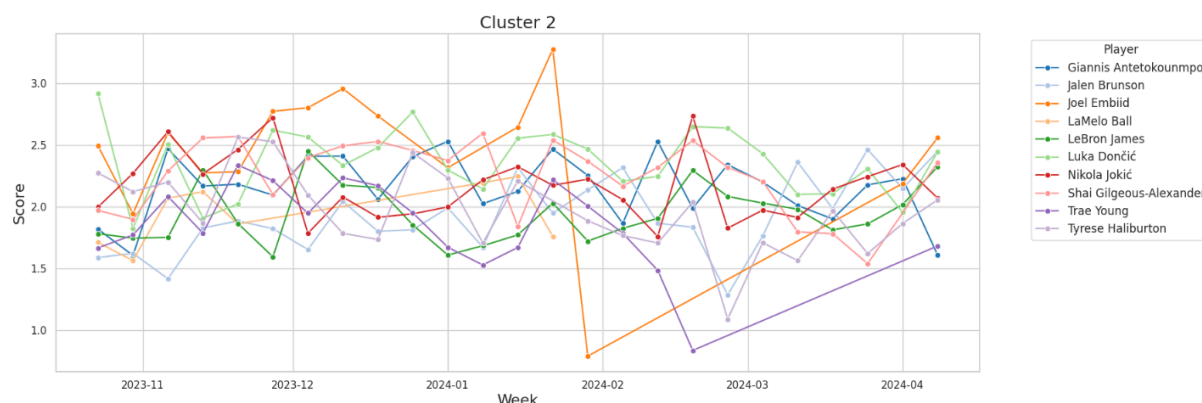


Figura 17 - Evolução do Top 10 jogadores estrela ao longo da época regular

No gráfico da Figura 18 destacam-se atletas como Anthony Davis, Victor Wembanyama e Domantas Sabonis, posicionados entre os melhores em termos de Score ao longo da temporada. Estes jogadores combinam uma significativa produção ofensiva com um impacto defensivo notável, justificando a sua classificação nesse grupo. O destaque vai para Victor Wembanyama, que arrecadou o prémio de Rookie of the Year e que apresentou um crescimento acentuado entre janeiro e março. Este desempenho colocou-o entre os três nomeados para o prémio de Defensive Player of the Year, terminando em segundo lugar na corrida.

Tal como já foi referido, é importante destacar a ausência do vencedor do prémio de Defensive Player of the Year, Rudy Gobert, entre os melhores jogadores do cluster 3. Como jogador menos orientado para a vertente ofensiva, considera-se natural que o Score de Rudy Gobert não seja tão elevado quanto os de outros atletas deste cluster.

Na NBA, a defesa é amplamente reconhecida como um aspecto crucial do jogo, muitas vezes determinante para o sucesso das equipas. Por isso, prémios como o de Defensive Player of the Year não avaliam apenas as estatísticas, mas também outros fatores intangíveis, sendo frequentemente alvo de debate devido às prestações defensivas de outros nomeados de elevado nível. Um dos aspetos relevantes na atribuição deste prémio é a prestação defensiva da equipa do jogador ao longo da época. Os Timberwolves, equipa de Rudy Gobert, destacou-se como a melhor equipa na vertente defensiva, liderando em métricas como a menor percentagem de eficácia dos adversários e o menor número de lançamentos tentados pelos oponentes. Estes números reforçam os argumentos a favor de Gobert no debate sobre quem merece o prémio de Melhor Defensor.

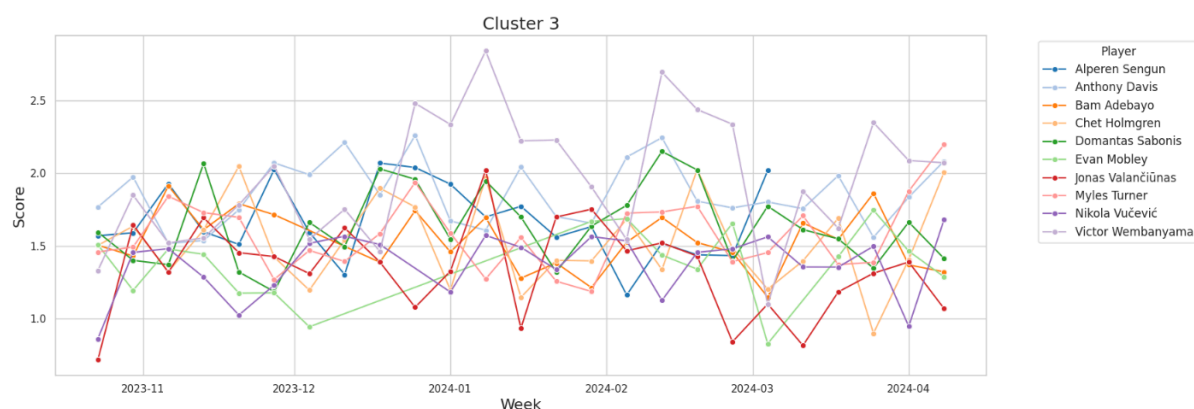


Figura 18 - Evolução do Top 10 jogadores defensivos ao longo da época regular

Com base na análise do gráfico da Figura 19, é possível observar tendências interessantes no Cluster 4, de role players ofensivos. Cam Whitmore e Jalen Smith destacaram-se inicialmente com scores elevados e consistentes nas primeiras semanas, indicando uma boa produção ofensiva. No entanto, com a progressão natural da época foram baixando o seu rendimento. Esta redução de rendimento fez com que Cam Whitmore perdesse o seu papel na rotação dos Rockets no início da temporada 2024/25.

Outro destaque importante é para Malik Monk, cujo desempenho consistente o levou a ser nomeado para o prémio Kia NBA Sixth Man of the Year. Embora não tenha vencido, a nomeação reflete a sua importância como role player ofensivo. Este exemplo reforça a valorização crescente da componente ofensiva de jogadores de banco na NBA moderna, capazes de manter a competitividade da equipa enquanto os titulares descansam.

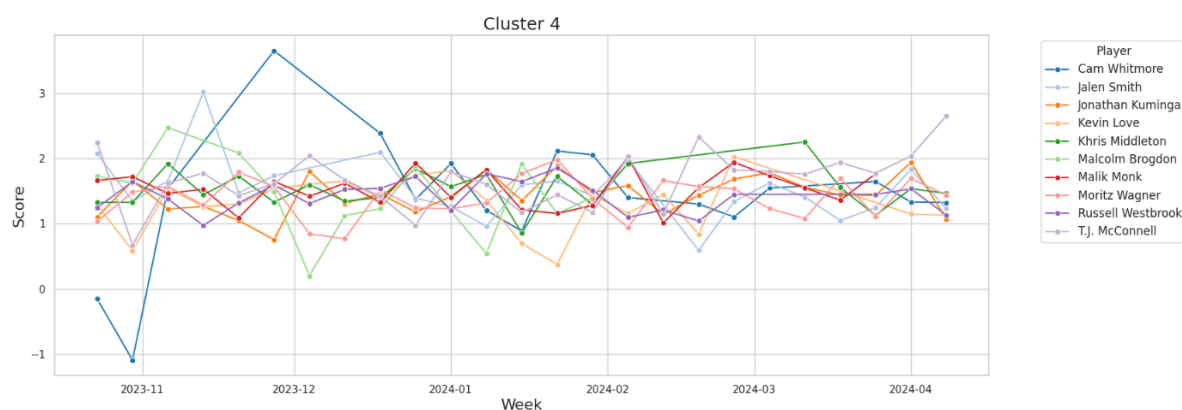


Figura 19 - Evolução do Top 10 role players ofensivos ao longo da época regular

O gráfico da Figura 20 apresenta a evolução semanal dos jogadores que saem do banco. Apesar de não serem protagonistas, estes jogadores asseguram estabilidade e suporte durante a temporada, com desempenhos consistentes refletidos em Scores uniformes ao longo do tempo. Esta regularidade permite aos treinadores contar com contribuições fiáveis, independentemente das circunstâncias do jogo.

O cluster reúne atletas de várias equipas, como o português Neemias Queta, Bones Hyland e Eugene Omoruyi, destacando a universalidade deste papel. Embora o desempenho geral seja estável, alguns jogadores registam picos em momentos específicos, frequentemente associados a maior tempo de jogo ou à substituição de titulares por motivo de lesão ou ausência. Casos como os de Neemias Queta e Eugene Omoruyi mostram aumentos significativos de contribuição em períodos pontuais, sublinhando a capacidade destes jogadores para capitalizar oportunidades de mais minutos em campo.

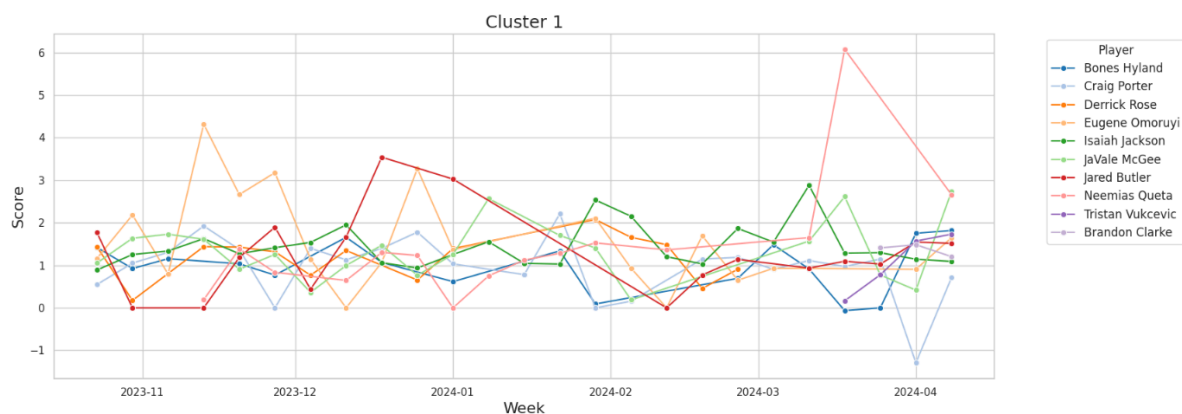


Figura 20 - Evolução do Top 10 bench players ao longo da época regular

Com base na análise do gráfico da Figura 21, observamos o Cluster 0, que inclui jogadores em início ou fim de carreira. Os seus scores relativamente elevados em certos momentos podem ser atribuídos a períodos limitados de tempo em campo, muitas vezes em jogos já decididos ou com menor competitividade. O padrão de evolução mensal dos scores é irregular: jogadores como Leonard Miller e Luka Garza apresentam picos de desempenho, enquanto outros mostram contribuições mais estáveis, refletindo a natureza das oportunidades de jogo e a gestão dos minutos.

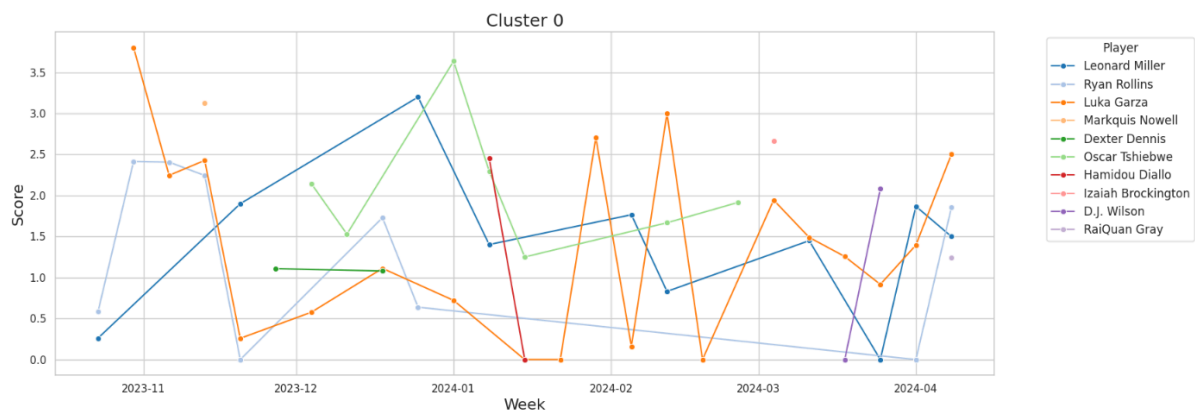


Figura 21 - Evolução do Top 10 jogadores menos ativos ao longo da época regular

8. CONCLUSÃO

A análise dos clusters revela importantes conclusões sobre a dinâmica dos diferentes grupos de jogadores na NBA. No Cluster de Jogadores Estrela, atletas como Joel Embiid, Luka Dončić e Giannis Antetokounmpo apresentam desempenhos consistentes, mas a atribuição do prêmio de MVP nem sempre reflete os números. Isto sugere que fatores intangíveis, como liderança, impacto global e o sucesso coletivo da equipa, são determinantes na escolha do MVP.

A análise do grupo de Jogadores Defensivos, inclui nomes como Victor Wembanyama, Anthony Davis e Domantas Sabonis, mas deixa de fora o vencedor do prêmio de Defensive Player of the Year, Rudy Gobert, que apesar do seu impacto defensivo, o seu Score não reflete esse domínio, o que indica que, para avaliar a defesa, as estatísticas ofensivas não têm qualquer tipo de influência/enviesamento da análise.

Nos Clusters de Role Players Ofensivos e Bench Players, observa-se a crescente importância dos jogadores de banco na NBA moderna. Estes jogadores contribuem significativamente para o equilíbrio da equipa, tanto na defesa como no ataque, e ajudam a garantir a continuidade do desempenho da equipa durante todo o jogo. A análise dos bench players e role players ofensivos demonstra que, numa liga cada vez mais orientada para o coletivo, o sucesso das equipas está intimamente ligado com a profundidade do plantel e da capacidade dos jogadores de banco em desempenharem papéis cruciais.

Estes clusters demonstram que a avaliação do desempenho de um jogador na NBA deve ser multifacetada, levando em conta o impacto direto no jogo, o papel que desempenha dentro da equipa e os fatores intangíveis que influenciam as decisões dos prêmios. Enquanto as estatísticas podem oferecer uma visão quantitativa, a verdadeira contribuição de um jogador para o sucesso da sua equipa vai além dos números e envolve aspectos como liderança, consistência ao longo da temporada e a capacidade de influenciar o jogo de maneira positiva. Portanto, as análises devem ter em conta estas múltiplas camadas, refletindo a complexidade do basquetebol moderno e o papel de cada jogador dentro da dinâmica das suas equipas.

9. LIMITAÇÕES E TRABALHOS FUTUROS

Este projeto, apesar de proporcionar uma visão abrangente do desempenho dos jogadores da NBA durante a época 2023/24, apresenta algumas limitações. Em primeiro lugar, a ausência de informações sobre as posições específicas dos jogadores limita a análise de padrões relacionados ao papel desempenhado em campo, o que poderia enriquecer a segmentação em clusters. Além disso, as análises focaram-se apenas em estatísticas quantitativas, sem incorporar fatores qualitativos e

intangíveis, como as contribuições/características estratégicas não refletidas diretamente nos números. Estas limitações impactam a capacidade de obter uma visão ainda mais detalhada e precisa sobre o desempenho dos atletas.

Como extensão deste projeto, trabalhos futuros poderiam incluir dados adicionais de outras fontes confiáveis para complementar as estatísticas e trazer maior diversidade à análise. Além disso, poderia incorporar-se informações sobre as posições dos jogadores e contexto de jogo, como a qualidade das equipas adversárias ou situações específicas de jogo, aprofundar o desenvolvimento de métricas personalizadas que levem em consideração mais fatores qualitativos e intangíveis, como impacto defensivo e liderança em campo. Por último, mas não menos importante, análises focadas em diferentes fases da época, como os playoffs, poderiam fornecer uma visão mais detalhada sobre o desempenho em contextos de maior pressão e competitividade.

BIBLIOGRAFIA

Zhang, S., Lorenzo, A., Gómez, M.-A., Mateus, N., Gonçalves, B., & Sampaio, J. (2018). Clustering performances in the NBA according to players' anthropometric attributes and playing experience. *Journal of Sports Sciences*, 36(13), 1506–1513.

Patel, R. (2017). *Clustering professional basketball players by performance* (Master's thesis, University of California, Los Angeles).

Richardson, C. A. (2019, December 19). *Evolution of a player: Transitions in NBA player classifications*. SSRN. <https://doi.org/10.2139/ssrn.3515711>

Muniz, M., & Flamand, T. (2022). A weighted network clustering approach in the NBA. *Journal of Sports Analytics*, 8(4), 251-275. <https://doi.org/10.3233/JSA-220584>

Papageorgiou, G., Sarlis, V., & Tjortjis, C. (2024). Evaluating the effectiveness of machine learning models for performance forecasting in basketball: A comparative study. *Knowledge and Information Systems*, 66(4333–4375). <https://doi.org/10.1007/s10115-024-02092-9>

Partnow, S. (2021). *The Midrange Theory*. United Kingdom: Triumph Books.

Dončić, L. (2023, May 22). *Luka Dončić explained why it's easier to score in the NBA than in EuroLeague*. BasketNews. <https://basketnews.com/news-168319-luka-doncic-explained-why-its-easier-to-score-in-the-nba-than-in-euroleague.html>

Herring, C. (2024). *NBA Awards 2024-2025: MVP, Rookie, Defensive Player, Coach*. ESPN. https://www.espn.com/nba/insider/story/_/id/41936258/nba-awards-2024-2025-mvp-rookie-defensive-player-coach

Basketball Reference. (2024). *NBA Awards 2024*. Basketball-Reference.com. https://www.basketball-reference.com/awards/awards_2024.html

Basketball Reference. (2023). *Daily leaders*. <https://www.basketball-reference.com/friv/dailyleaders.fcgi?month=12&day=01&year=2023&type=all>