

CLUSTERS DE PERFORMANCE NA NBA

Projeto Aplicado II

Grupo 16

2024/2025

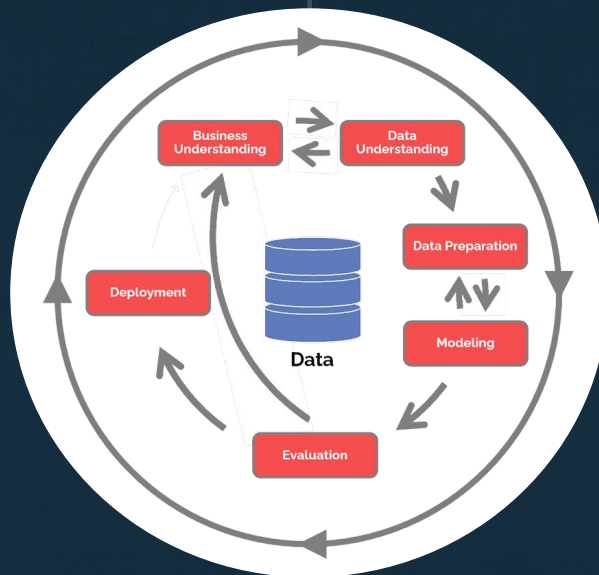


INTRODUÇÃO

INTRODUÇÃO E FRAMEWORK METODOLÓGICO

OBJETIVO

- #1** Analisar a performance dos jogadores da NBA durante a época regular de 2023-2024, identificando padrões e agrupando-os em clusters com características semelhantes.



PORQUÊ?

- #2** Que estatísticas contribuem para diferenciar mais os jogadores?
- #3** Que jogadores tiveram maior impacto ofensivo e defensivo? Como foi a progressão ao longo da época?
- #4** A atribuição das NBA Awards (como o MVP) foi justa? O que nos dizem os dados sobre esta classificação?
- #5** Que equipas partilham características comuns?

ÉPOCA REGULAR

PLAYOFFS

Outubro 2023

Abril 2024

Junho 2024

30 EQUIPAS

15 na Conferência Este
15 na Conferência Oeste

16 EQUIPAS

8 na Conferência Este
8 na Conferência Oeste

**PLAY-
IN**

82 JOGOS

41 em casa
41 fora

**ELIMINAÇÃO
DIRETA**

01

02

03

04

05

06

07

DATA UNDERSTANDING

EXTRAÇÃO DO DATASET INICIAL

DATASET INICIAL

#1

Estatísticas diárias dos jogadores da NBA durante a temporada regular recolhidas através de web scraping com Selenium do website Basketball Reference (Daily Leaders)

PROCESSO DE EXTRAÇÃO

#1

Geração de URLs: Cada dia da época regular tem um URL único.

#2

Selenium: Automatização da navegação e extração com Selenium.

#3

Seletores XPath: Utilizados para localizar as tabelas de estatísticas diárias no código HTML.

#4

Agregação: Os dados extraídos diariamente são consolidados num só DataFrame.

DICIONÁRIO DE VARIÁVEIS

Informações do jogador e do jogo

- **Player:** Nome do jogador.
- **Tm:** Equipa à qual o jogador pertence.
- **vs:** Indica se a equipa do jogador jogou em casa ou fora.
- **Opp:** Equipa adversária enfrentada.
- **W/L:** Resultado do jogo (vitória ou derrota).
- **Date:** Data do jogo.

Informações de desempenho

- **FG, FGA, FG%:** Cestos de campo convertidos, tentativas e percentagem de acerto.
- **3P, 3PA, 3P%:** Cestos de 3 pontos convertidos, tentativas e percentagem de acerto.
- **FT, FTA, FT%:** Lançamentos livres convertidos, tentativas e percentagem de acerto.
- **PTS:** Pontos totais marcados.
- **ORB, DRB, TRB:** Ressaltos ofensivos, defensivos e totais.
- **+/-:** Diferença no marcador enquanto o jogador esteve em campo.
- **MP:** Minutos jogados.
- **AST:** Assistências.
- **STL:** Roubos de bola.
- **BLK:** Desarmes de lançamento.
- **TOV:** Perdas de bola.
- **PF:** Faltas cometidas.

ANÁLISE EXPLORATÓRIA INICIAL



01

02

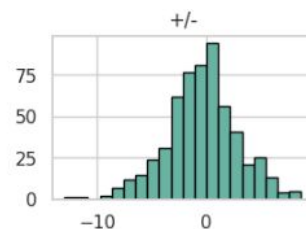
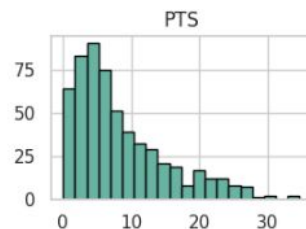
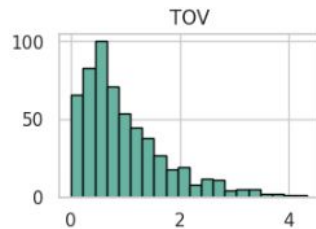
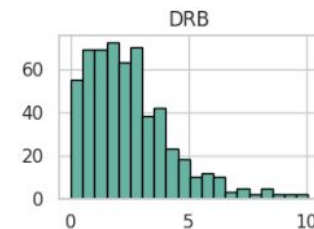
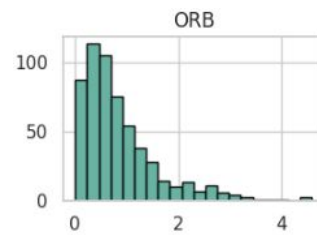
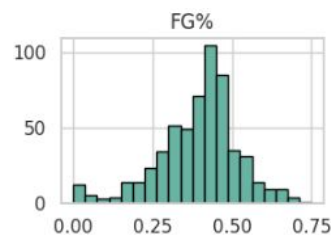
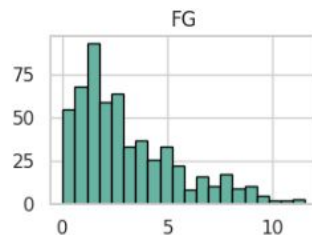
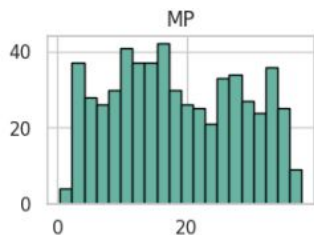
03

04

05

06

07



* para facilitar a análise exploratória inicial foi feito algum pré-processamento dos dados, 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

ANÁLISE EXPLORATÓRIA INICIAL

01

#1

A correlação elevada entre turnovers e assistências (0.85) indica que jogadores mais ativos na criação de jogadas, com maior posse de bola, também são mais suscetíveis a cometer erros.

02

03

04

05

#2

A correlação mais baixa entre ressaltos ofensivos e pontos (0.34) reforça a mudança no estilo de jogo da NBA, com maior foco em lançamentos de três pontos, ao invés de ressaltos ofensivos e put-backs.

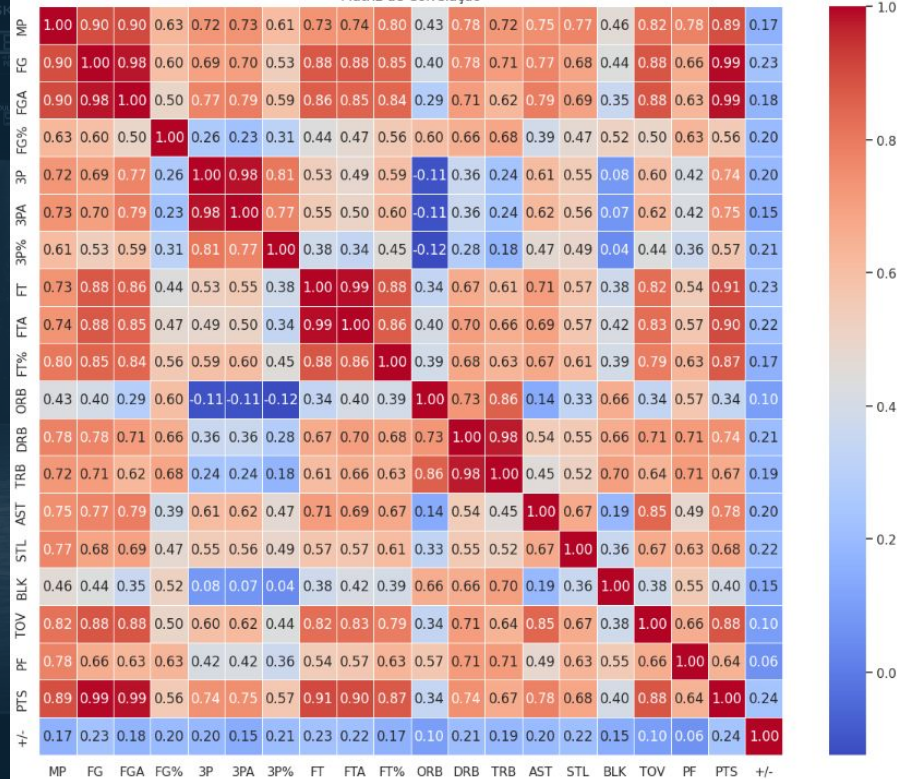
06

07

#3

A variável +/- apresenta uma distribuição simétrica em torno de zero, indicando um equilíbrio entre os jogadores com impacto positivo e negativo no marcador durante os jogos.

Matriz de Correlação



DATA PREPARATION

ETAPAS REALIZADAS

- 01
- 02
- 03
- 04**
- 05
- 06
- 07

PROCESSAMENTO E TRATAMENTO DE VALORES OMISSOS

#1

Com base nas estatísticas diárias de cada jogador, gerar um novo DataFrame com as estatísticas médias da época.

#2

Imputação nos valores omissos de percentagens de lançamento (FG%, 3P%, FT%) com 0, para jogadores que não realizaram tentativas de lançamento.

#3

Para a variável +/-, foram imputados os valores omissos com 0 para jogadores sem minutos jogados (MP = 0) e com a mediana dos valores de +/- para os restantes jogadores já que existiam apenas 44 valores omissos.

**FEATURE ENGINEERING E TRANSFORMAÇÃO DE VARIÁVEIS**

#1

MP (minutos jogados) foi convertida de formato MM:SS para formato decimal.

#2

Criação de novas variáveis para contar número de jogos em casa, número de vitórias, adversário contra quem marcou mais pontos, ressaltos e assistências.

#3

Criação de uma fórmula de Score que atribui mais pesos aos PTS, AST, BLK e STL, menos peso aos TBR, pontos negativos para TOV, estando tudo normalizado pelos MP.

OUTLIERS

#1

Mantidos no dataset, pois representam dados reais e importantes para a validade das conclusões.

MODELING

ANÁLISE DE COMPONENTES PRINCIPAIS

01

02

03

04

05

06

07

#1

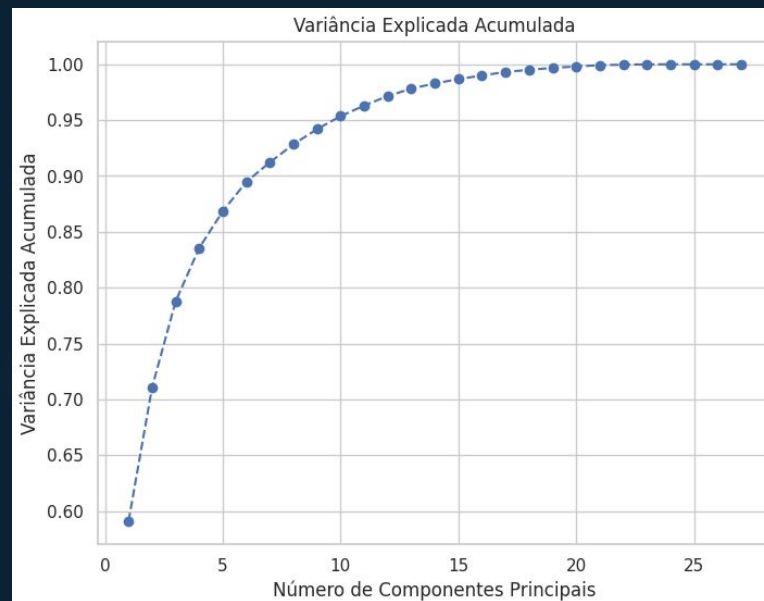
A análise de componentes principais (PCA) foi aplicada após a padronização das variáveis numéricas.

#2

A PCA identificou que cinco componentes principais explicam aproximadamente 86,8% da variância total dos dados, com base na análise da variância explicada.

#3

O critério de Kaiser, que seleciona componentes com eigenvalues maiores que 1, sugeriu quatro componentes, mas optou-se por manter cinco componentes para uma explicação mais robusta.



CARACTERIZAÇÃO DAS COMPONENTES PRINCIPAIS

ANÁLISE DOS LOADINGS

- 01
- 02
- 03
- 04
- 05**
- 06
- 07

PC1

Relacionada com a produção ofensiva, capturando tanto a eficiência quanto o volume de jogo ofensivo, com variáveis como minutos jogados (MP), arremessos convertidos (FG) e pontos (PTS).

PC2

Reflete o desempenho defensivo e o controlo do garrafão, associando-se a variáveis como ressaltos ofensivos (ORB), defensivos (DRB) e bloqueios (BLK).

PC3

Está ligada ao desempenho em jogos em casa vs fora, considerando variáveis como o número de jogos em casa (Home_Games_Count) e vitórias em casa (Home_Wins_Count).

PC4

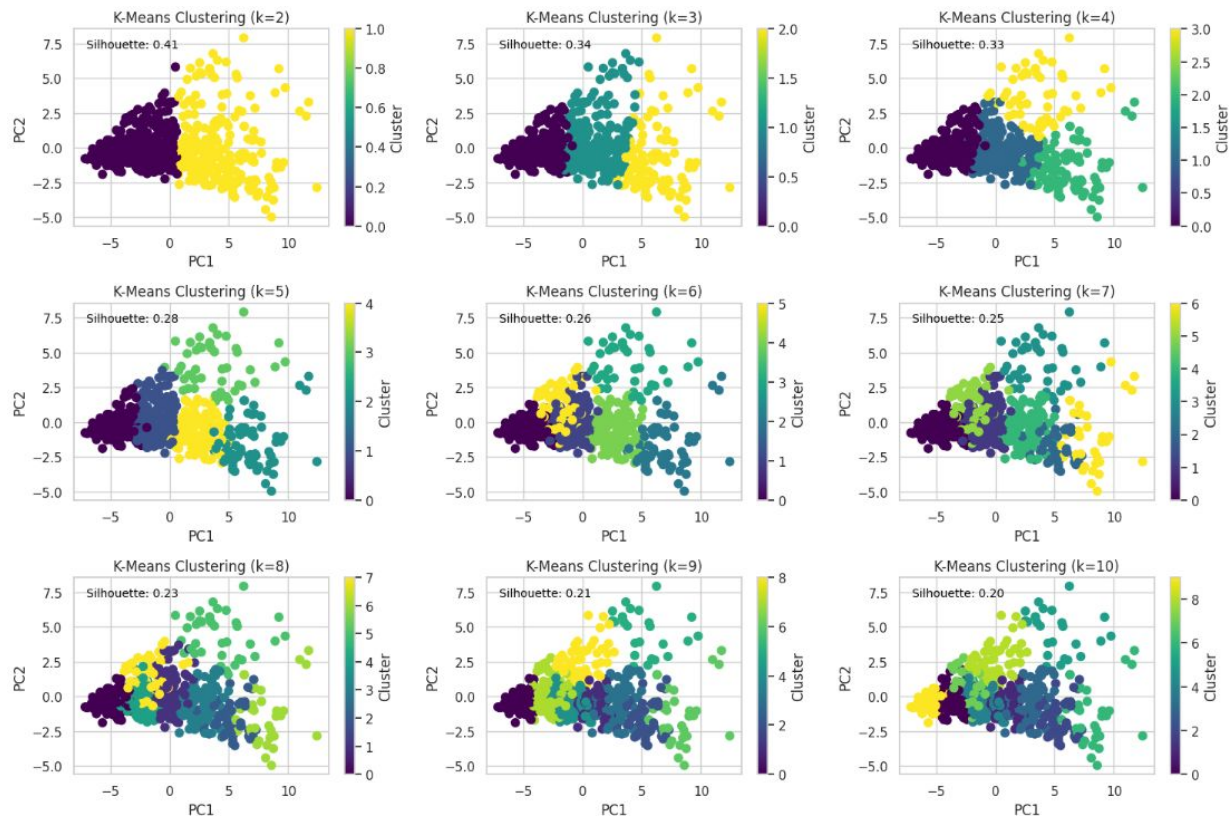
Mede o impacto geral de um jogador no jogo, sendo influenciada pela diferença de pontos (+/-) e vitórias/derrotas em casa, refletindo a contribuição global do jogador.

PC5

Combina eficiência no ataque (arremessos de 3 pontos, FT%) e habilidades técnicas.

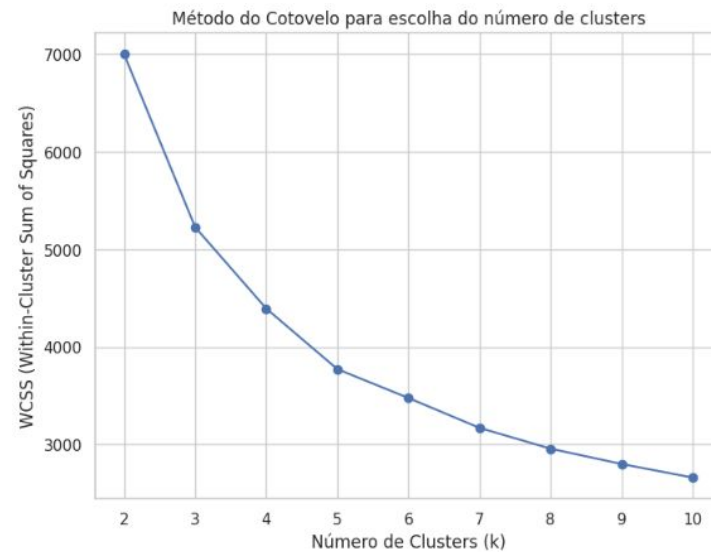
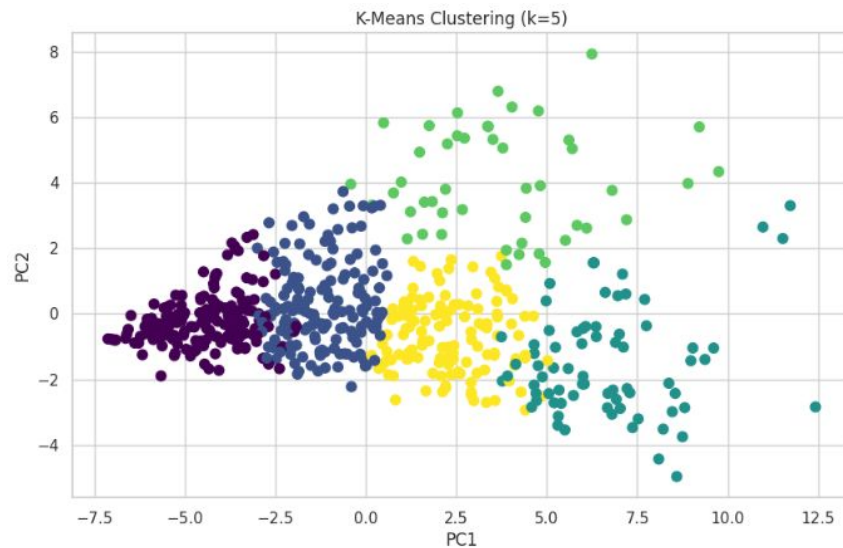
K-MEANS CLUSTERING

- 01
- 02
- 03
- 04
- 05
- 06**
- 07

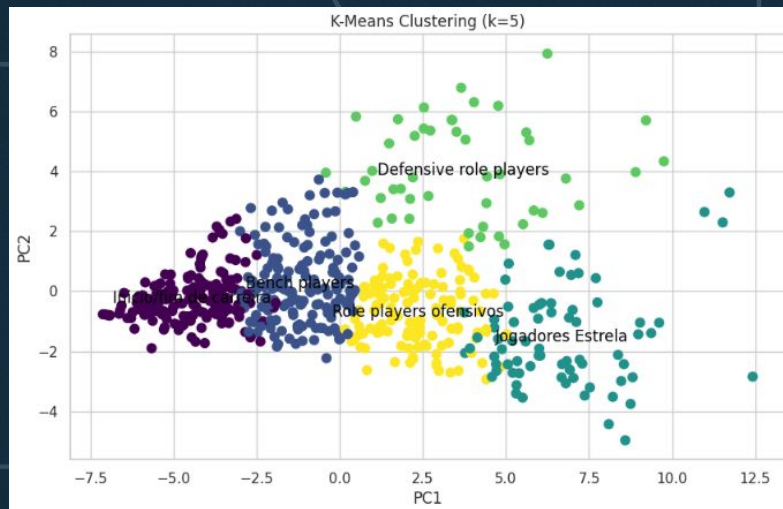


K-MEANS CLUSTERING

- 01
- 02
- 03
- 04
- 05
- 06
- 07



EVALUATION



- **Cluster 0 ("Jogadores Menos Ativos / Início ou Fim de Carreira"):** Baixa média de minutos (7.75), com baixo desempenho ofensivo e defensivo.
- **Cluster 1 ("Jogadores de Papel Secundário/Bench Players"):** Média de 15.69 minutos, baixo volume ofensivo e boa contribuição em ressaltos e assistências.
- **Cluster 2 ("Jogadores Estrela"):** Alta média de minutos (33.56), grande destaque ofensivo (21.98 pontos) e bom desempenho defensivo.
- **Cluster 3 ("Defensive Role Players"):** Média de 26.91 minutos e forte contribuição defensiva, especialmente em ressaltos.
- **Cluster 4 ("Role Players Ofensivos"):** Média de 25.75 minutos e boa produção ofensiva (11.28 pontos), com percentagens de lançamento próximas do cluster 2.

01

02

03

04

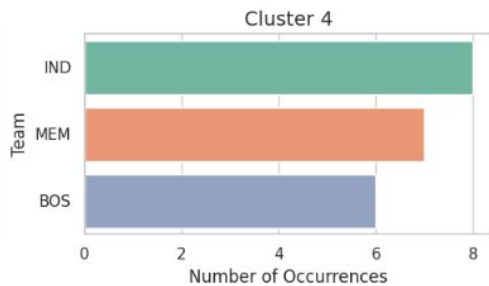
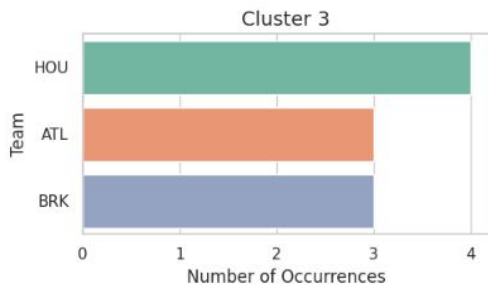
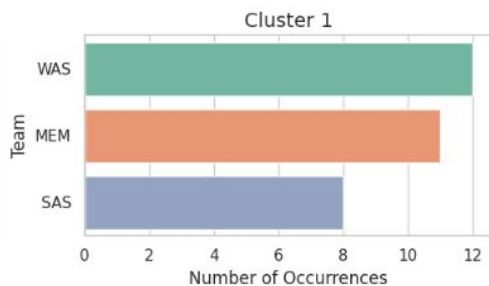
05

06

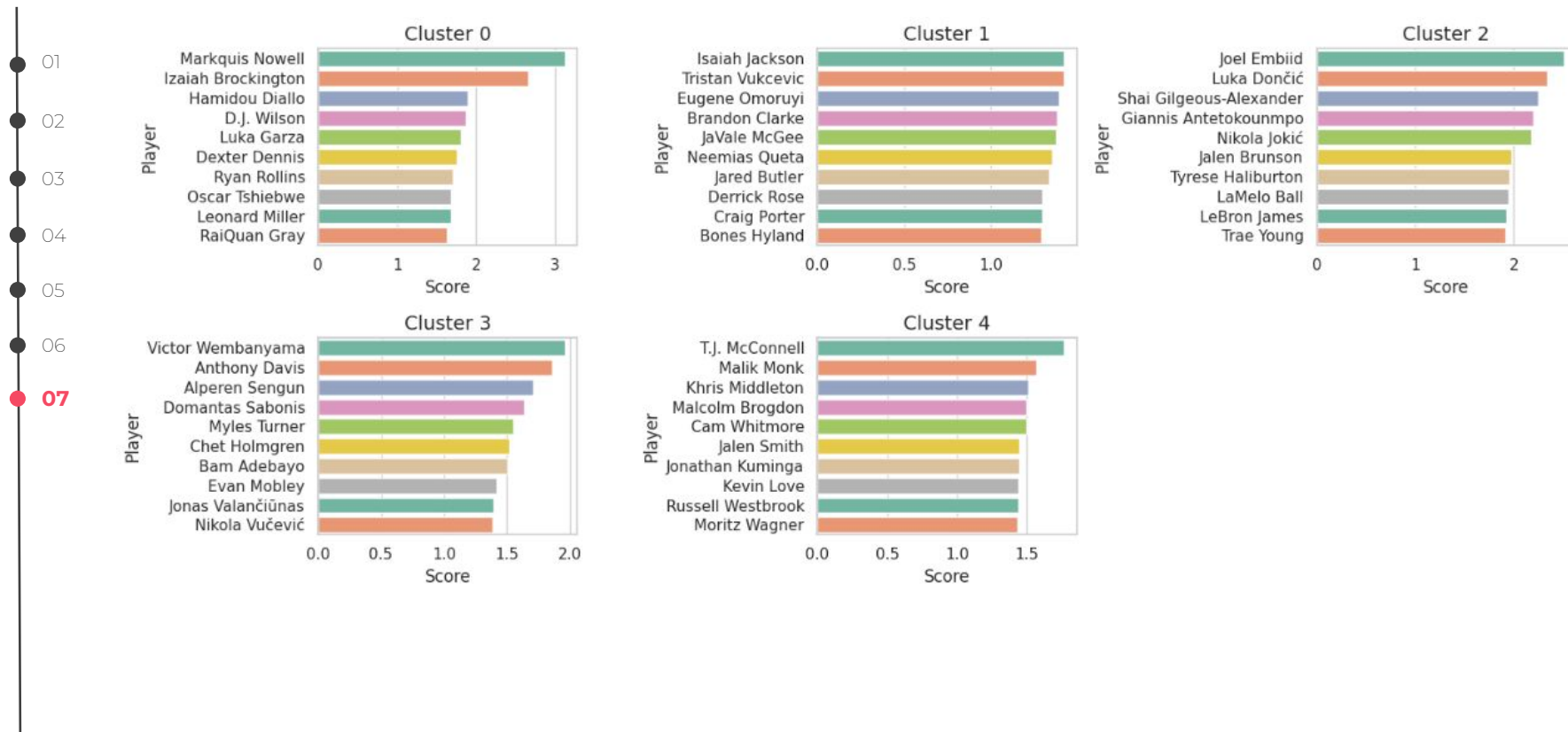
07

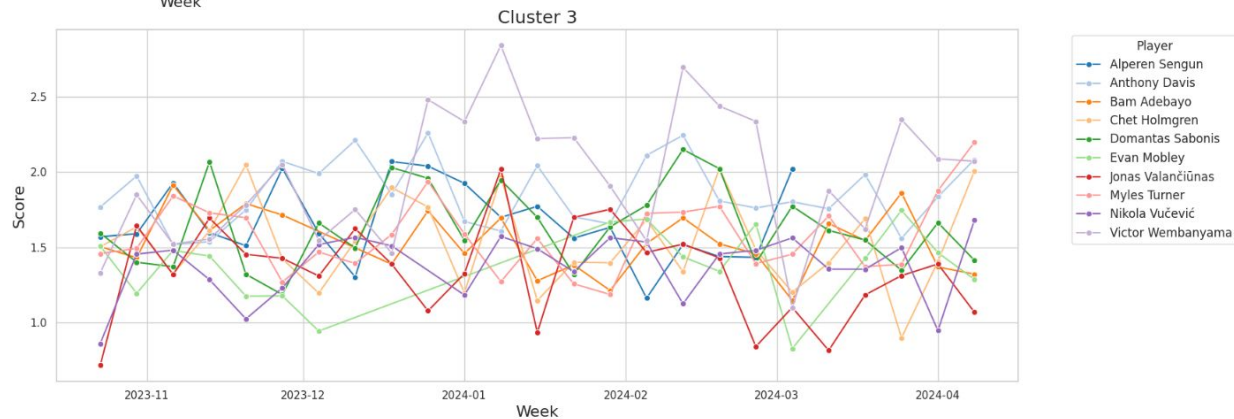
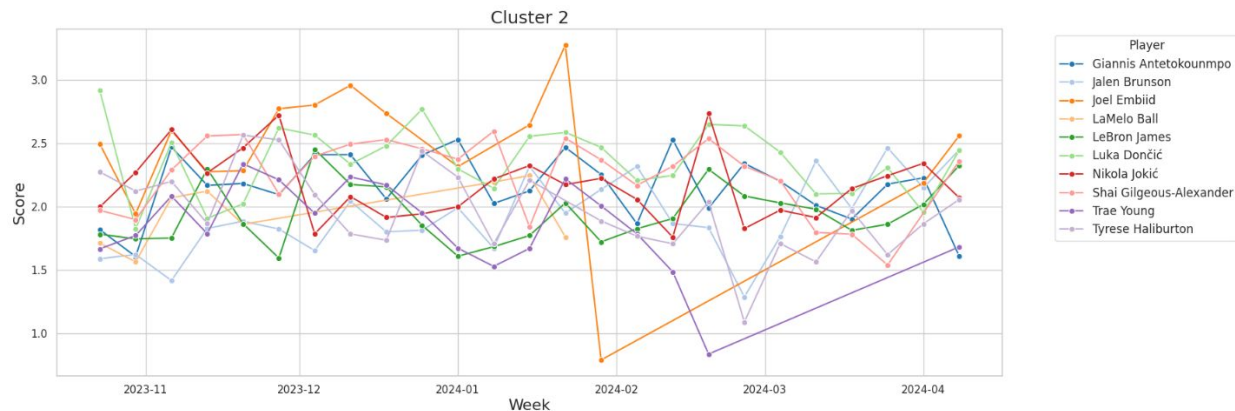
K-MEANS CLUSTERING

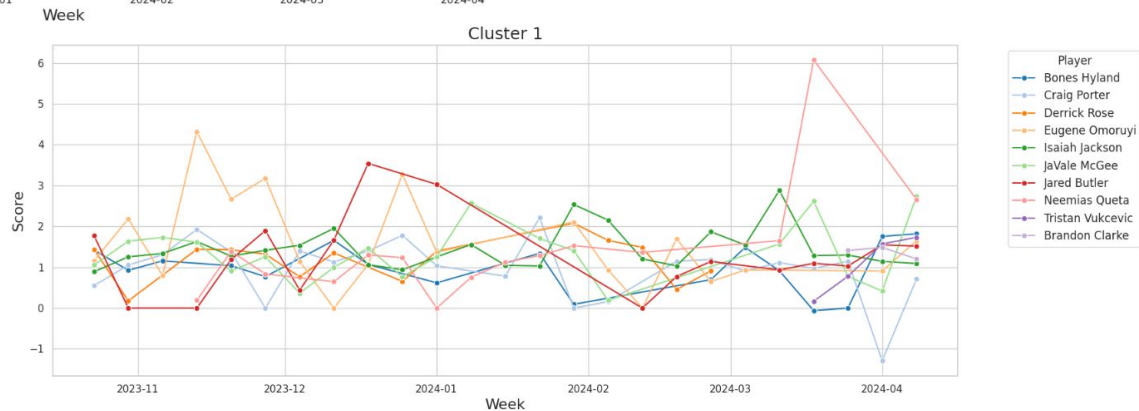
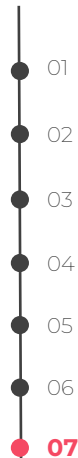
- 01
- 02
- 03
- 04
- 05
- 06
- 07

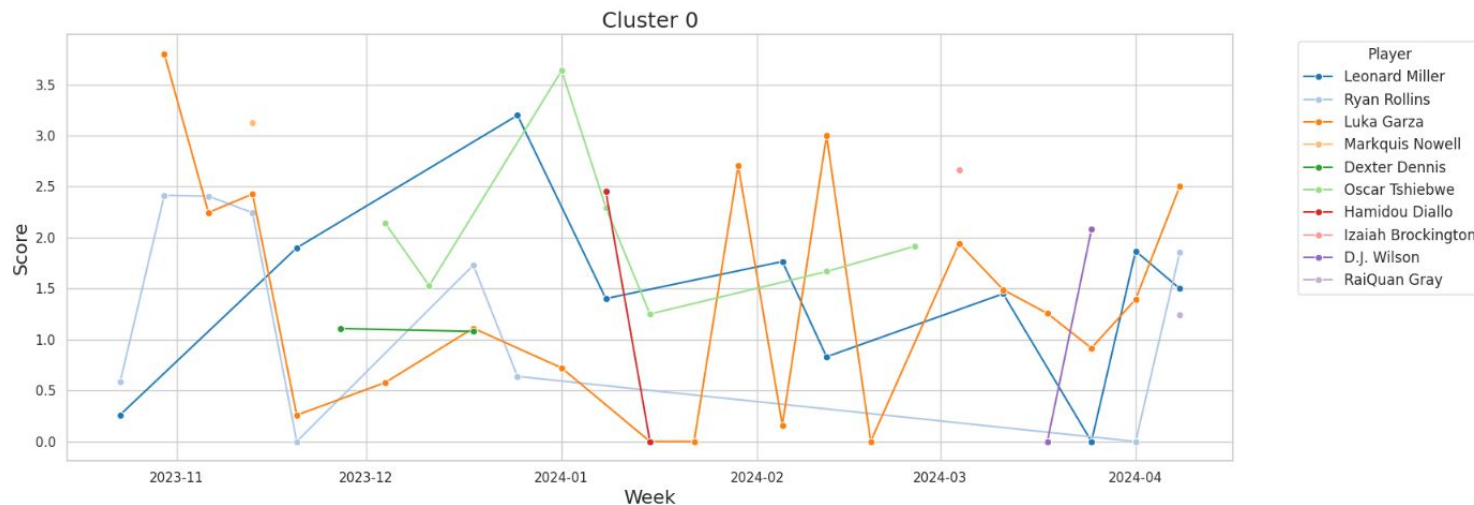


K-MEANS CLUSTERING









CONCLUSÕES, LIMITAÇÕES E TRABALHOS FUTUROS

- 01 **#1** Jogadores como **Joel Embiid**, **Luka Dončić** e **Antetokounmpo** apresentam desempenhos consistentes, mas o **MVP não reflete apenas números**, considerando **intangíveis** como liderança e impacto coletivo.
- 02
- 03 **#2** Rudy **Gobert**, premiado com **Defensive Player of the Year**, não se encontra entre os melhores jogadores do cluster defensivo, pela sua fraca capacidade ofensiva, o que sugere que **a atribuição do prémio não é influenciada por fatores ofensivos**.
- 04
- 05 **#3** A importância crescente dos jogadores de banco, que equilibram a equipa tanto na defesa como no ataque, reflete a evolução para um jogo coletivo, onde a **profundidade do plantel é crucial para o sucesso**.
- 06
- 07 **#4** A maior proximidade entre os clusters de jogadores estrela e jogadores ofensivos nas análises de clusters reflete as características do **estilo de jogo da NBA**, que **favorece o ataque** e a liberdade de movimento.
- #5** A falta de dados sobre **posições** e **fatores qualitativos** limita a análise. Futuros trabalhos poderiam incluir dados adicionais, análise do impacto defensivo, liderança e contexto de jogo, além de avaliar o **desempenho nos playoffs**.

OBRIGADO!