



UNIVERSIDADE DO SUL DE SANTA CATARINA

A3: ANÁLISE DE DADOS DA NBA

Florianópolis

2024

EDUARDO ALVARENGA
EDUARDO DOS SANTOS ARAÚJO
MARCOS OLIVEIRA
FÁBIO SÉRGIO SANTOS FILHO
THIAGO GENTIL

A3: ANÁLISE DE DADOS DA NBA

Professor: Jorge Werner

Florianópolis

2024

SUMÁRIO

1	INTRODUÇÃO.....	3
2	DESENVOLVIMENTO	4
2.1	CRONOGRAMA.....	4
2.2	ESCOPO E OBJETIVO	5
2.3	COLETA E PREPARAÇÃO DE DADOS	6
2.4	TECNOLOGIA	6
2.4.1	Estrutura de Dados	6
2.4.2	Processo de Trabalho.....	7
2.5	EXPLORAÇÃO INICIAL DE DADOS	8
2.6	ANÁLISE ESTATÍSTICA DESCRITIVA	8
2.7	SELEÇÃO DE TÉCNICAS DE ANÁLISE	9
2.8	COMUNICAÇÃO DOS RESULTADOS	9
3	CONCLUSÃO	12
4	ARTEFATOS	13

1 INTRODUÇÃO

Este projeto se dedica à análise de dados temporais da National Basketball Association (NBA), concentrando-se no desempenho dos jogadores ao longo de um período específico. Utilizando técnicas avançadas de análise de dados, como análise de séries temporais, por exemplo, e ferramentas como JClick, Excel e outras, este estudo tem como objetivo explorar padrões e tendências nos dados para informar estratégias táticas, avaliações de desempenho individual e previsões de resultados de jogos.

Durante o desenvolvimento deste trabalho, os dados foram coletados, tratados e analisados meticulosamente para garantir a precisão e a relevância das conclusões. A integração de diferentes fontes de dados, como o Kaggle e o site oficial da NBA, proporcionou uma base robusta para as análises realizadas. A análise exploratória dos dados revelou insights valiosos sobre o impacto das variáveis como minutos de jogo, eficiência de arremessos e contribuições defensivas no desempenho geral dos jogadores e equipes.

Este trabalho não apenas visa aprofundar o entendimento do desempenho na NBA, mas também tem potencial para fornecer informações estratégicas úteis para treinadores, analistas de desempenho, comentaristas esportivos e outras partes interessadas.

2 DESENVOLVIMENTO

2.1 CRONOGRAMA

Cronograma e Distribuição de Tarefas

Brainstorming e Planejamento

- **Eduardo Araujo:** Fornecer a ideia principal do projeto relacionado à NBA, sugerir quais dados seriam interessantes trazer para a apresentação e como esses dados poderiam ser usados.
- **Eduardo Alvarenga:** Propor a utilização do Kaggle como fonte de dados.
- **Thiago:** Oferecer dicas e insights sobre NBA com base em seu conhecimento.
- **Marcos:** Apoiar nas ideias e colaborar com sugestões.

Coleta de Dados

- **Eduardo Alvarenga:** Buscar e baixar os dados relevantes do Kaggle, realizar o tratamento dos dados e a pré-análise.
- **Fábio:** Organizar os dados coletados em tabelas para facilitar a análise.
- **Eduardo Araujo:** Validar os dados com base em seu conhecimento da NBA e pesquisas acerca do tema, garantindo que sejam relevantes e precisos.
- **Marcos e Thiago:** Apoiar no processo de coleta e organização dos dados, oferecendo suporte onde necessário.

Análise de Dados e Visualização

- **Eduardo Araujo:** Definir como os dados serão utilizados na apresentação, identificar quais visualizações serão mais impactantes e quem será o público-alvo da apresentação.
- **Fábio:** Criar visualizações de dados utilizando ferramentas apropriadas (como Qlik ou outras), baseando-se nas diretrizes fornecidas por Eduardo Araujo.
- **Eduardo Alvarenga:** Oferecer feedback sobre as visualizações e sugerir possíveis melhorias.
- **Marcos:** Apoiar no desenvolvimento das visualizações, colaborando com ideias e sugestões.

Preparação Final

- **Eduardo Araujo:** Estruturar a apresentação, integrando as visualizações de dados e explicações sobre como eles se encaixam no contexto da NBA. Certificar-se de que as visualizações de dados (gráficos e tabelas) estejam bem posicionadas e acompanhadas de explicações claras sobre sua relevância e interpretação.
- **Fábio:** Finalizar as tabelas e visualizações, garantindo que estejam prontas para a apresentação. Polir os gráficos e tabelas, corrigindo possíveis erros e melhorando a estética. Assegurar que todas as visualizações sejam claras, legíveis e prontas para serem integradas.
- **Eduardo Alvarenga:** Garantia da credibilidade das fontes, com foco especial no uso do Kaggle.
- **Thiago:** Estar pronto para fornecer respostas detalhadas e contextualizadas sobre a liga, as estatísticas dos jogadores e a relevância dos dados apresentados.
- **Marcos:** Apoiar na preparação geral da apresentação, revisando o conteúdo e sugerindo ajustes finais.

2.2 ESCOPO E OBJETIVO

Este projeto visa realizar uma análise detalhada dos dados temporais da National Basketball Association (NBA), concentrando-se no desempenho de jogadores ao longo das diversas temporadas da liga. Utilizando um conjunto abrangente de dados estatísticos, o objetivo principal é extrair insights significativos que possam beneficiar diferentes áreas de interesse.

O foco central deste projeto reside em explorar como as estatísticas individuais e coletivas dos jogadores da NBA podem ser analisadas para melhorar diversos aspectos:

Prevenção de Lesões: Investigar a correlação entre o desempenho dos jogadores e o risco de lesões físicas, considerando variáveis como faltas pessoais e estatísticas defensivas, visando promover práticas de jogo mais seguras e estratégias de treinamento mais eficazes.

Análise Técnica e Estratégica: Identificar padrões de desempenho que possam ser utilizados para aprimorar estratégias de equipe, como alocação de tempo de jogo e seleção de formações, com base em dados históricos robustos.

Previsões Esportivas: Desenvolver modelos preditivos que utilizem estatísticas passadas para prever resultados de jogos, incorporando variáveis como o desempenho individual dos jogadores, histórico de confrontos entre equipes e localização das partidas.

Análise de Séries Temporais: Analisar a evolução do desempenho individual e coletivo dos jogadores ao longo das temporadas, identificando tendências que possam oferecer insights valiosos para estratégias de longo prazo e análises de tendências.

Algumas perguntas relevantes acerca do tema, podem ser elencadas abaixo:

- Quais jogadores apresentam uma maior contribuição de pontos em toda sua carreira?
- Qual a quantidade de pontos, assistências e bloqueios médio por temporada?
- Quais posições cometem o maior número de faltas na liga e que jogadores lideram estes numeros?

2.3 COLETA E PREPARAÇÃO DE DADOS

Os dados foram coletados a partir de datasets da plataforma Kaggle, contendo informações relevantes de cada temporada da NBA, como:

- **Estatísticas de Jogadores:** pontos, assistências, rebotes, tocos, roubos de bola, porcentagem de arremessos de quadra, porcentagens de arremessos de 3 pontos, porcentagem de lances livres, minutos jogados, jogos disputados, etc.

Para o tratamento de dados utilizou-se de Python para remover as linhas duplicadas e realizar agrupamento de certos dados.

2.4 TECNOLOGIA

2.4.1 Estrutura de Dados

- **Fontes de Dados:**
 - **Kaggle:** Utilização de conjuntos de dados disponíveis para download relacionados à NBA, bem como estatísticas de jogadores.
 - **Site da NBA:** Utilização para validação de dados e obtenção de informações adicionais sobre estatísticas, jogos, e outros dados relevantes.
- **Infraestrutura Necessária:**
 - **Hardware:** Computadores com capacidade suficiente para suportar o Qlik Sense e Excel, garantindo capacidade de processamento e armazenamento adequados para manipulação de grandes conjuntos de dados.
 - **Conectividade:** Acesso estável à internet para download de dados do Kaggle e acesso ao site da NBA.
- **Software e Plataformas**
 - **Qlik Sense:** Plataforma de BI e visualização de dados utilizada para criar dashboards interativos, analisar tendências e padrões nos dados da NBA.

- **Jupyter Notebook e Python:** Utilizado para o pré-processamento inicial dos dados, limpeza, transformação e manipulação. Também utilizado para a análise descritiva.
- **Kaggle:** Fonte de dados para conjuntos específicos relacionados à NBA, proporcionando acesso a uma ampla gama de dados históricos e atuais.
- **Site da NBA:** Referência para validação de dados, garantindo a precisão das estatísticas e informações utilizadas no projeto.
- **Linguagens Utilizadas**
 - **Qlik Script:** Linguagem de script específica do Qlik Sense utilizada para carregar, transformar e modelar dados.
 - **Python:** Utilizada para manipulação e limpeza de dados.

2.4.2 Processo de Trabalho

Coleta de Dados: Download de conjuntos de dados relevantes do Kaggle, incluindo estatísticas de jogadores, equipes e resultados de jogos da NBA.

Validação de Dados: Verificação da consistência e precisão dos dados por meio do site oficial da NBA para garantir a confiabilidade das informações utilizadas.

Pré-processamento:

- **Limpeza Inicial:** Remoção de dados duplicados, preenchimento de valores faltantes e padronização de formatos de dados.
- **Tratamento de Dados de Jogadores:** Identificação e tratamento de casos onde jogadores tenham registros múltiplos devido a mudanças de time durante a mesma temporada.
 - Exemplo: Alguns jogadores podem aparecer mais de uma vez nos dados devido a transferências entre times. É necessário unificar esses registros para evitar conflitos e assegurar que todas as estatísticas sejam consolidadas corretamente.

Importação para Qlik Sense: Carregamento dos dados preparados e tratados no Qlik Sense para análise e visualização.

Análise e Visualização: Utilização das funcionalidades do Qlik Sense para criar dashboards interativos que permitam explorar padrões, tendências e insights nos dados da NBA.

2.5 EXPLORAÇÃO INICIAL DE DADOS

A exploração inicial dos dados envolveu:

- **Análise descritiva:** Foram calculadas estatísticas como média, mediana, mínimo e máximo para variáveis relevantes ao longo de cada temporada. Por exemplo, a média de pontos por jogador, o número total de assistências e o número de bloqueios realizados.
- **Visualização dos dados:** Foram criados gráficos para visualizar a distribuição das variáveis e identificar possíveis relações entre elas. Isso incluiu gráficos de barras para comparar jogadores com mais pontos, gráficos de linha para mostrar tendências ao longo das temporadas e gráficos de dispersão para analisar correlações entre diferentes métricas.

2.6 ANÁLISE ESTATÍSTICA DESCRITIVA

Buscou-se desenvolver mais detalhadamente a análise descritiva já feita na exploração inicial dos dados, através de:

- **Análise de correlação:** cálculo de coeficientes de correlação para identificar relações lineares entre variáveis. (Encontra-se no arquivo `matriz_de_correlacao.csv`). Foi retirado as seguintes informações:

pts e ast (0.71): Há uma forte correlação positiva entre pontos e assistências. Isso significa que, em geral, jogadores que pontuam mais tendem a fazer mais assistências.

pts e trb (0.72): Há uma forte correlação positiva entre pontos e rebotes. Jogadores que pontuam mais também tendem a pegar mais rebotes.

pts e blk (0.37): Há uma correlação moderada positiva entre pontos e bloqueios. Embora não tão forte quanto as outras, ainda sugere que jogadores que pontuam mais também fazem mais bloqueios.

ast e trb (0.38): Há uma correlação moderada positiva entre assistências e rebotes.

ast e blk (0.12): Há uma correlação fraca positiva entre assistências e bloqueios.

trb e blk (0.56): Há uma correlação forte positiva entre rebotes e bloqueios, sugerindo que jogadores que pegam muitos rebotes também tendem a fazer muitos bloqueios.

2.7 SELEÇÃO DE TÉCNICAS DE ANÁLISE

Técnica Escolhida: Análise de séries temporais

A análise de séries temporais foi a técnica escolhida para este estudo, com o objetivo de examinar as estatísticas dos jogadores da NBA ao longo de toda a história da liga. Essa abordagem nos permitiu identificar tendências, padrões e anomalias nos dados, fornecendo insights valiosos sobre a evolução do desempenho dos atletas e como as diferentes eras da NBA podem ter influenciado suas estatísticas.

A análise de séries temporais é particularmente adequada para este contexto, pois os dados das estatísticas dos jogadores da NBA são naturalmente sequenciais e dependentes do tempo. Ao considerar a ordem temporal dos dados, podemos observar como as estatísticas evoluíram ao longo dos anos, identificar pontos de inflexão e avaliar o impacto de eventos específicos na liga.

2.8 COMUNICAÇÃO DOS RESULTADOS

1. Introdução e Contexto do Projeto

Este relatório apresenta uma análise detalhada dos dados da NBA com o objetivo de fornecer insights para diversas aplicações, incluindo análise de desempenho, estudos da NBA e transmissões esportivas. Utilizamos técnicas estatísticas para explorar padrões e tendências nos dados coletados de fontes confiáveis como o Kaggle e o site oficial da NBA.

2. Metodologia Utilizada

Os dados foram coletados e preparados utilizando ferramentas como Qlik Sense e Excel para análise exploratória. A metodologia incluiu a aplicação de técnicas de regressão para modelagem preditiva de desempenho de jogadores, agrupamento para segmentação de dados relevantes, e análise de séries temporais para identificar tendências ao longo das temporadas.

3. Principais Resultados

- **Posições por Temporada:** Analisar a distribuição de jogadores por posição ao longo das temporadas permite identificar tendências na composição das equipes. Por exemplo, se houve uma mudança de preferência por determinadas posições em períodos de tempo específicos, isso pode refletir adaptações estratégicas das equipes para se ajustar a novos estilos de jogo ou mudanças nas regras da liga.
- **Pontos, Assistências e Bloqueios:** Avaliar os pontos, assistências e bloqueios por jogo por posição fornece insights valiosos sobre a contribuição específica de cada posição para o desempenho geral da equipe. Por exemplo, se certas posições têm

consistentemente mais assistências, isso pode indicar uma ênfase na criação de jogadas ofensivas a partir dessas posições.

- **Faltas Cometidas por Posição:** Identificar quais posições cometem mais faltas pode ser crucial para desenvolver estratégias defensivas mais disciplinadas. Isso pode incluir ajustes nas técnicas de marcação ou na distribuição de minutos de jogo para minimizar o impacto das faltas em momentos cruciais das partidas
- **Aproveitamento de 2 e 3 pontos:** Analisar o aproveitamento de arremessos de 2 pontos e 3 pontos por posição ajuda a entender a eficácia dos jogadores em diferentes áreas do campo. Isso é fundamental para ajustar as estratégias de jogo, como determinar quem deve ter mais liberdade para arremessar de longa distância ou quem deve focar mais em jogadas próximas à cesta.
- **Aproveitamento de Tiros Livres:** O aproveitamento de lances livres por posição pode indicar quais jogadores são mais confiáveis sob pressão e quais posições podem precisar de mais treinamento nesse aspecto fundamental do jogo. Melhorar o aproveitamento de tiros livres pode resultar em mais pontos convertidos e influenciar diretamente o resultado das partidas.

4. Recomendações e Próximos Passos

Com base nos resultados obtidos, recomendamos algumas estratégias práticas para maximizar o desempenho e a segurança dos jogadores na NBA:

Para começar, é fundamental otimizar as estratégias táticas da equipe utilizando os insights derivados da análise. Isso inclui ajustar abordagens de jogo para aumentar a eficiência em diferentes contextos competitivos.

Além disso, sugerimos expandir a análise incorporando variáveis adicionais, como a idade dos jogadores e as condições específicas das partidas. Isso não só enriquece a compreensão dos padrões de desempenho, mas também melhora a precisão das previsões futuras.

Por fim, é essencial manter os modelos de análise atualizados regularmente com novos dados. Isso garante que as previsões permaneçam relevantes e precisas ao longo do tempo, permitindo ajustes contínuos e informados nas estratégias da equipe.

Essas recomendações não apenas visam otimizar o desempenho esportivo, mas também são fundamentais para promover um ambiente de jogo seguro e eficaz na NBA. Além disso, podem ser aplicadas de forma abrangente em análises de desempenho, transmissões esportivas

e outras áreas que dependem de insights precisos e atualizados para melhorar a compreensão e a estratégia no basquete profissional.

3 CONCLUSÃO

A análise dos dados revela insights valiosos que podem ser aplicados estrategicamente no contexto da NBA. Utilizando informações detalhadas sobre o desempenho dos jogadores, como aproveitamento de arremessos, bloqueios e faltas cometidas, é possível aprimorar táticas e estratégias de jogo de maneira mais eficiente. Isso não apenas maximiza as habilidades específicas dos jogadores, mas também fundamenta decisões estratégicas dos treinadores com base em dados concretos de desempenho.

Além disso, há uma correlação evidente entre o número de faltas pessoais e o risco aumentado de lesões físicas entre os jogadores da NBA. Jogadores que frequentemente se envolvem em situações defensivas intensas, como bloqueios e desarmes agressivos, enfrentam maior probabilidade de lesões musculares, articulares e traumáticas. Compreender essa relação oferece uma oportunidade crucial para a prevenção e gestão eficaz da saúde dos atletas, destacando a importância contínua de análises detalhadas para identificar padrões e tendências que possam impactar o bem-estar e o desempenho dos jogadores ao longo da temporada.

Por fim, análises detalhadas dos dados não apenas impactam diretamente o desempenho e a saúde dos jogadores, mas também enriquecem a narrativa durante as transmissões esportivas. Ao oferecer insights baseados em métricas como aproveitamento de arremessos e eficiência defensiva, os comentaristas podem proporcionar uma análise profunda e contextualizada durante os jogos. Isso não só eleva a experiência de assistir aos jogos, mas também demonstra como a análise de dados é uma ferramenta poderosa para engajar e educar o público sobre os aspectos técnicos e táticos do basquete profissional.

4 ARTEFATOS

Player_Totals: Dataset inicial

Player Totals Updated: Dataset final utilizado após a preparação dos dados:

Dashboards_1 e Dashboards_2: Imagens referente aos resultados finais da análise

Estatísticas_descritivas_por_temporada: Arquivo csv contendo apenas as análises estatísticas feitas no dataset.

Matriz de correlação: Arquivo csv contendo apenas os resultados da criação da matriz de correlação

NBA_Stats: Arquivo Jupyter Notebook contendo o código final do processo de Tratamento de Dados e Pré-análise.