

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO (PUC-RIO)  
CURSO DE PÓS GRADUAÇÃO EM CIÊNCIA DE DADOS E ANALYTICS**

**EDUARDO CAMARGO RIBEIRO BORGES**

**ANÁLISE DE DADOS HISTÓRICOS DA NBA: CONSTRUÇÃO DE UM PIPELINE ETL  
COM SNOWFLAKE E VISUALIZAÇÃO DE DADOS UTILIZANDO QLIK**

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>3</b>
1.1	Justificativa	3
<b>2</b>	<b>DEFINIÇÃO DO PROBLEMA E OBJETIVOS</b>	<b>4</b>
2.1	Objetivo Geral	4
2.2	Perguntas de Análise	4
<b>3</b>	<b>METODOLOGIA</b>	<b>5</b>
3.1	Plataformas utilizadas	5
3.1.1	Snowflake	5
3.1.2	Qlik Cloud	5
3.2	Fonte de dados	5
3.2.1	NBA Database	5
3.2.2	NBA Players stats since 1950	6
3.3	Modelagem proposta e Catálogo de Dados	7
3.3.1	Modelagem	7
3.3.1.1	Dimensões	8
3.3.1.2	Fatos	8
3.3.2	Catálogo de Dados	9
<b>4</b>	<b>PIPELINE DE DADOS E ETL - SNOWFLAKE</b>	<b>13</b>
4.1	Landing Zone - Staging Data	13
4.2	Curated Zone - Tables	15
4.3	Curated Zone - Data Processing	16
4.4	Consumption Zone - Staging Data	18
<b>5</b>	<b>ETL EXTRA E PAINÉIS - QLIK CLOUD</b>	<b>23</b>
5.1	Conexão com o Snowflake	23
5.2	Carregando dados	23
5.2.1	Extração	23
5.2.2	Transformação	24
5.2.3	Carregamento	28
5.3	Criação do painel BI	29
5.3.1	Dashboard	30
5.3.2	Times	30
5.3.3	Jogadores	32
5.3.4	Estatísticas Jogadores	33
5.3.5	Comparativo Jogadores	34
5.3.6	Draft	35
<b>6</b>	<b>ANÁLISES, RESULTADOS OBTIDOS E CONCLUSÕES</b>	<b>37</b>

<b>6.1</b>	<b>Respostas . . . . .</b>	<b>37</b>
6.1.1	Quais as tendências e padrões em Drafts da NBA? . . . . .	37
6.1.2	Existe um impacto da altura dos jogadores no desempenho em quadra? . . . . .	38
6.1.3	Como as estatísticas dos jogadores evoluíram ao longo do tempo? . . . . .	39
6.1.4	Quais equipes tiveram maior impacto na revelação de grandes estrelas? . . . . .	40
<b>6.2</b>	<b>Análise da qualidade da base de dados utilizada . . . . .</b>	<b>41</b>
<b>7</b>	<b>CONCLUSÃO GERAL DO PROJETO . . . . .</b>	<b>43</b>
<b>8</b>	<b>AUTOAVALIAÇÃO . . . . .</b>	<b>44</b>

## **1 INTRODUÇÃO**

O uso de análise de dados tem se tornado essencial em diversas áreas, incluindo o esporte. No basquete profissional, a NBA (“National Basketball Association”) é uma das ligas que mais investem em análise estatística para tomada de decisões, influenciando desde a seleção de jogadores até o desenvolvimento de estratégias de jogo. Este trabalho explora a utilização de ferramentas de Business Intelligence (BI) para analisar dados históricos da NBA, identificando tendências e padrões a partir de estatísticas de jogadores e times.

### **1.1 Justificativa**

A NBA gera um grande volume de dados históricos sobre jogadores, times e jogos, mas muitos desses dados estão dispersos e difíceis de analisar. Este trabalho justifica-se pela necessidade de organizar e analisar esses dados para extrair insights valiosos sobre o desempenho dos jogadores e das equipes. Utilizando o Snowflake para processamento e armazenamento dos dados e o Qlik Sense para visualização, buscamos transformar dados brutos em informações acessíveis e interativas. A combinação dessas ferramentas oferece uma solução eficiente para compreender tendências e padrões na NBA, contribuindo para uma análise mais profunda da liga e suas temporadas.

## **2 DEFINIÇÃO DO PROBLEMA E OBJETIVOS**

O objetivo principal é construir um pipeline de dados eficiente para coletar, transformar e visualizar estatísticas da NBA, possibilitando a análise de tendências e extração de insights. A plataforma BI utilizada para apresentação dos dados será o Qlik, e o Snowflake será empregado para processamento e armazenamento dos dados (ETL).

### **2.1 Objetivo Geral**

Como a análise de estatísticas históricas da NBA pode auxiliar na identificação de padrões e tendências para tomada de decisões dentro do esporte?

### **2.2 Perguntas de Análise**

- a) Quais as tendências e padrões em Drafts da NBA? Esta pergunta analisa como as escolhas no Draft mudaram ao longo dos anos, focando em fatores como a posição dos jogadores, a origem universitária e a altura.
- b) Existe um impacto da altura dos jogadores no desempenho em quadra? Aqui, busca-se entender se existe uma correlação entre a altura dos jogadores e seu desempenho em estatísticas chave, como pontos, rebotes e assistências.
- c) Como as estatísticas dos jogadores evoluíram ao longo do tempo? Essa questão examina a evolução das principais estatísticas individuais, como pontos e eficiência de arremesso, ao longo das décadas.
- d) Quais equipes tiveram maior impacto na revelação de grandes estrelas? Investiga quais equipes tiveram mais sucesso em desenvolver ou revelar grandes jogadores, seja através de escolhas de draft ou programas de desenvolvimento.

### 3 METODOLOGIA

Neste capítulo, será apresentada a metodologia adotada no trabalho, abrangendo as plataformas utilizadas, as fontes de dados, o catálogo detalhado dos dados analisados e, por fim, o detalhamento do dashboard BI.

#### 3.1 Plataformas utilizadas

##### 3.1.1 Snowflake

O Snowflake foi utilizado para a ingestão, processamento e armazenamento dos dados no pipeline ETL, desempenhando um papel fundamental na organização e estruturação dos dados. A escolha do Snowflake se deu devido a várias características vantajosas da plataforma, incluindo seu período de testes de 30 dias, que permitiu validar a solução sem gastos. A plataforma também conta com uma série de ferramentas integradas para a construção de pipelines de dados, simplificando o processo de transformação e carregamento dos dados. Sua facilidade de integração com outras plataformas, como Qlik, garantiu uma transição fluida dos dados para as análises e visualizações, otimizando todo o fluxo de trabalho e permitindo uma exploração eficiente das informações.

##### 3.1.2 Qlik Cloud

O Qlik Cloud foi utilizado para a visualização e exploração dos dados e para a criação de dashboards BI dinâmicos. A plataforma oferece diversas opções gráficas e ferramentas analíticas, permitindo a criação de visualizações avançadas, como gráficos interativos, painéis de controle e relatórios personalizados. Além disso, a plataforma oferece um período de testes, permitindo a execução deste trabalho sem custos.

#### 3.2 Fonte de dados

Após uma pesquisa, na plataforma Kaggle, a qual possui um repositório com diversos *datasets*, foram encontrados dois *datasets*.

##### 3.2.1 NBA Database

O primeiro *dataset* chamado NBA Database, criado por Wyatt Walsh<sup>1</sup>, o *dataset* possui diversos dados da NBA, como por exemplo, dados sobre jogadores, times, *drafts*, jogos, etc... Os dados são de até 2023 e foram coletados pelo autor

---

<sup>1</sup> NBA Database Disponível em <https://www.kaggle.com/datasets/wyattowalsh/basketball> acessado em: 03/2025

utilizando a API do site stats.nba<sup>2</sup>, que contém dados históricos informados pela própria NBA.

No total existem 16 tabelas com diferentes informações sobre a NBA, mas para este projeto foram utilizadas apenas 8, sendo elas:

- a) **common\_player\_info**: Tabela com informações detalhadas dos jogadores;
- b) **draft\_combine\_stats**: Tabela com dados de estatísticas do *draft combine* dos jogadores;
- c) **draft\_history**: Tabela com informações de *drafts* realizados de cada jogador;
- d) **game**: Tabela de jogos, com detalhes estatísticos e times que jogaram cada jogo.
- e) **officials**: Tabela de arbitros que atuaram nos jogos;
- f) **player**: Tabela de todos os jogadores registrados na NBA;
- g) **team**: Tabela com informações principais dos times que atuam na NBA;
- h) **team\_details**: Tabela com informações detalhadas dos times;

As demais tabelas não foram relevantes para a aplicação deste trabalho e, portanto, não serão detalhadas neste documento.

### 3.2.2 NBA Players stats since 1950

O segundo *dataset* chamado NBA Players stats since 1950, criado por Omri Goldstein<sup>3</sup>, o *dataset* possui dados relacionados especificamente com dados estatísticos de jogadores e suas estatísticas por temporada. Os dados são de até 2017 e foram coletados pelo autor utilizando dados do site Basketball Reference<sup>4</sup>, que contém dados históricos coletados pela plataforma.

No total existem 3 tabelas com dados dos jogadores, mas para este projeto foram utilizadas apenas 1, sendo ela:

- a) **season\_stats**: Tabela com dados estatísticos dos jogadores de todas as temporadas até 2017;

Essa tabela sera relacionada aos dados do outro dataset para realização das análises por temporada.

<sup>2</sup> stats.nba Disponível em <https://www.nba.com/stats> acessado em: 03/2025

<sup>3</sup> NBA Players stats since 1950 Disponível em <https://www.kaggle.com/datasets/drgilermo/nba-players-stats> acessado em: 03/2025

<sup>4</sup> Basketball Reference Disponível em <https://www.basketball-reference.com> acessado em: 03/2025

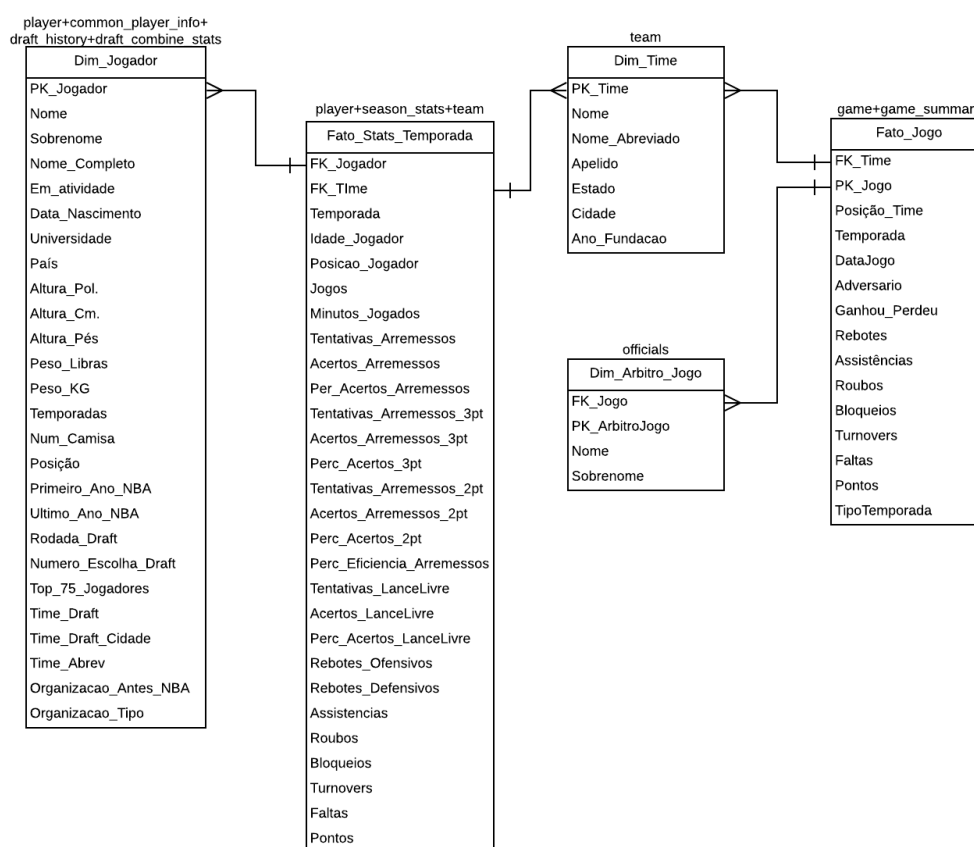
### 3.3 Modelagem proposta e Catálogo de Dados

#### 3.3.1 Modelagem

A modelagem dos dados seguiu o padrão *Snowflake Schema* contendo 3 dimensões e 2 fatos, a modelagem inicial foi desenvolvida utilizando a plataforma LucidChart<sup>5</sup> e demonstrado na figura 1.

A escolha pela utilização do esquema Snowflake se deu pela necessidade de maior flexibilidade na modelagem do data warehouse, uma vez que foram utilizadas duas tabelas fato distintas: Fato\_Stats\_Temporada e Fato\_Jogo. Essas tabelas representam granularidades diferentes — a primeira está organizada por temporada (ano), enquanto a segunda é estruturada por jogo (data específica). Além disso, cada uma relata um tipo diferente de informação: uma focada no desempenho acumulado dos jogadores ao longo da temporada, e a outra nos detalhes de cada partida individual. Dessa forma, a separação em duas tabelas fato se mostrou essencial para garantir consistência, clareza analítica e maior poder de exploração nos dashboards desenvolvidos.

**Figura 1 – Snowflake Schema**



<sup>5</sup> LucidChart Disponível em <https://www.lucidchart.com> acessado em: 03/2025



### 3.3.1.1 Dimensões

- a) **Dim\_Jogador:** Dimensão de jogadores e seus dados de *draft*

Tabelas dos datasets utilizadas para sua criação:

- *player*;
- *common\_player\_info*;
- *draft\_history*;
- *draft\_combine\_stats*.

- b) **Dim\_Time:** Dimensão dos times da NBA

Tabelas dos datasets utilizadas para sua criação:

- *team*.

- c) **Dim\_Arbitro\_Jogo:** Dimensão dos arbitros que atuaram em jogos da NBA

Tabelas dos datasets utilizadas para sua criação:

- *officials*.

### 3.3.1.2 Fatos

- a) **Fato\_Stats\_Temporada:** Fato com estatísticas acumuladas dos jogadores em cada temporada

Tabelas dos datasets utilizadas para sua criação:

- *player*;
- *season\_stats*;
- *team*.

Como a base dessa fato vem do *dataset* NBA Players stats since 1950, foi necessário juntar as tabelas *player* e *team* do outro *dataset*, criando chaves manualmente para manter a integridade dos dados.

- b) **Fato\_Jogo:** Fato com dados e estatísticas de todos os jogos dentro de cada temporada

Tabelas dos datasets utilizadas para sua criação:

- *game*;
- *game\_summary*;
- *team*.

### 3.3.2 Catálogo de Dados

A seguir, serão apresentados detalhadamente os dados de todas as dimensões e fatos apresentadas anteriormente na modelagem.

#### a) **Dim\_Jogador:**

1. **PK\_Jogador:** Chave primaria (Number, 12,0);
2. **Nome:** Primeiro nome do jogador (Varchar, 20);
3. **Sobrenome:** Sobrenome do jogador (Varchar, 30);
4. **Nome\_Completo:** Nome completo do jogador (Varchar, 40);
5. **Em\_Atividade:** Flag para definir se o jogador esteve em atividade até 2023 (Number, 1,0);
6. **Data\_Nascimento:** Data de nascimento do jogador (Date);
7. **Universidade:** Universidade em que o jogador jogou (Varchar, 80);
8. **País:** Pais de origem do jogador (Varchar, 40);
9. **Altura\_Pol.:** Altura do jogador em polegadas (Number, 3,0);
10. **Altura\_Cm:** Altura do jogador em centímetros (Number, 3,0);
11. **Altura\_Pés:** Altura do jogador em pés (Varchar, 4);
12. **Peso\_Libras:** Peso do jogador em Libras (Number, 3,0);
13. **Peso\_KG:** Peso do jogador em KG (Number, 3,0);
14. **Temporadas:** Temporadas as quais o jogador atuou até 2023 (Number, 3,0);
15. **Num\_Camisa:** Número da camisa do Jogador (Number, 3,0);
16. **Posição:** Posição do jogador (Varchar, 20);
17. **Primeiro\_Ano\_NBA:** Primeiro ano de NBA (Date);
18. **Ultimo\_Ano\_NBA:** Último ano de NBA registrado até 2023 (Date);
19. **Rodada\_Draft:** Rodada de escolha do jogador no draft (Varchar, 3);
20. **Numero\_Escolha\_Draft:** Número da escolha do jogador na rodada do draft (Varchar, 3);
21. **Top\_75\_Jogadores:** Flag para definir se o jogador é um dos maiores 75 de todos os tempos (Varchar, 1);
22. **Time\_Draft:** Time que escolheu o jogador no draft (Varchar, 80);
23. **Time\_Draft\_Cidade:** Cidade do time que escolheu o jogador (Varchar, 40);

24. **Time\_Abrev**: Nome abreviado do time que escolheu o jogador (Varchar, 3);
25. **Organizacao\_Antes\_NBA**: Organização a qual o jogador atuou antes de ser "draftado"(Varchar 80);
26. **Organizacao\_Tipo**: Tipo de organização a qual o jogador atuou antes de ser "draftado"(Varchar 18).

b) **Dim\_Time**:

1. **PK\_Time**: Chave primaria (Number, 12,0);
2. **Nome**: Nome do time (Varchar, 40);
3. **Nome\_Abreviado**: Nome abreviado do time (Varchar, 30);
4. **Apelido**: Apelido do time (Varchar, 30);
5. **Estado**: Estado do time (Varchar, 40);
6. **Cidade**: Cidade do time (Varchar, 40);
7. **Ano\_Fundacao**: Ano de Fundação do time (Date).

c) **Dim\_Arbitro\_Jogo**:

1. **PK\_ArbitroJogo**: Chave primaria (Number, 12,0);
2. **FK\_Jogo**: Chave estrangeira do jogo (Number, 12,0);
3. **Nome**: Nome do Arbitro (Varchar, 20);
4. **Sobrenome**: Sobrenome do Arbitro (Varchar, 40).

d) **Fato\_Stats\_Temporada**:

1. **FK\_Jogador**: Chave estrangeira do jogador (Number, 12,0);
2. **FK\_Time**: Chave estrangeira do time (Number, 12,0);
3. **Temporada**: Temporada em que o jogador atuou (Varchar, 20);
4. **Idade\_Jogador**: Idade do jogador (Varchar, 2,0);
5. **Posicao\_Jogador**: Posição do jogador (Varchar, 20);
6. **Jogos**: Quantidade de jogos em que o jogador atuou (Number, 3,0);
7. **Minutos\_Jogados**: Minutos jogados pelo jogador (Number, 5,0);
8. **Tentativas\_Aremessos**: Total de arremessos (Number, 4,0);
9. **Acertos\_Aremessos**: Total de acertos de arremessos (Number, 3,0);

10. **Perc\_Acertos\_Aremessos**: Percentual de acertos sobre arremessos (Number, 4,2);
11. **Tentativas\_Aremessos\_3pt**: Total de arremessos de 3pt (Number, 4,0);
12. **Acertos\_Aremessos\_3pt**: Total de acertos de arremessos de 3pt (Number, 4,0);
13. **Perc\_Acertos\_3pt**: Percentual de acertos sobre arremessos de 3pt (Number, 4,2);
14. **Tentativas\_Aremessos\_2pt**: Total de arremessos de 2pt (Number, 4,0);
15. **Acertos\_Aremessos\_2pt**: Total de acertos de arremessos de 2pt (Number, 4,0);
16. **Perc\_Acertos\_2pt**: Percentual de acertos sobre arremessos de 2pt (Number, 4,2);
17. **Perc\_Eficiencia\_Aremessos**: Percentual de arremessos eficazes, leva em consideração que os arremessos de 3pt valem mais (Number, 4,2);
18. **Tentativas\_LanceLivre**: Total de lances livres (Number, 4,0);
19. **Acertos\_LanceLivre**: Total de acertos em lances livres (Number, 4,0);
20. **Perc\_Acertos\_LanceLivre**: Percentual de acertos sobre lances livres (Number, 4,2);
21. **Rebotes\_Ofensivos**: Total de rebotes ofensivos (Number, 4,0);
22. **Rebotes\_Defensivos**: Total de rebotes defensivos (Number, 4,0);
23. **Assistencias**: Total de assistências (Number, 4,0);
24. **Roubos**: Total de roubos de bola (Number, 4,0);
25. **Bloqueios**: Total de bloqueios (Number, 4,0);
26. **Turnovers**: Total de turnovers (Number, 4,0);
27. **Faltas**: Total de faltas (Number, 4,0);
28. **Pontos**: Total de pontos (Number, 4,0).

e) **Fato\_Jogo**:

1. **PK\_Jogo**: Chave Primaria (Number, 12,0);
2. **FK\_Time**: Chave estrangeira do time (Number, 12,0);

3. **Posicao\_Time**: Define se o time jogou em casa ou fora (Varchar, 20);
4. **Temporada**: Temporada (Varchar, 20);
5. **Data\_Jogo**: Data do jogo (Date);
6. **Adversario**: Adversário (Varchar, 40);
7. **Ganhou\_Perdeu**: Define se o time venceu ou perdeu a partida (Varchar, 12);
8. **Rebotes**: Total de rebotes (Number, 4,0);
9. **Assistências**: Total de assistências (Number, 4,0);
10. **Roubos**: Total de roubos (Number, 4,0);
11. **Bloqueios**: Altura do jogador em polegadas (Number, 4,0);
12. **Turnovers**: Total de turnovers (Number, 4,0);
13. **Faltas**: Total de faltas (Number, 4,0);
14. **Pontos**: Total de pontos (Number, 4,0);
15. **Tipo\_Temporada**: Define qual é a situação da temporada (regular/-playoffs) (Varchar, 30).

Esses dados foram preparados para consumo no Qlik. No próximo capítulo, será detalhado o processo de estruturação do pipeline de dados, explicando as etapas necessárias para alcançar esse resultado.

## 4 PIPELINE DE DADOS E ETL - SNOWFLAKE

Conforme mencionado anteriormente, a criação do pipeline de dados foi realizada utilizando o software Snowflake, que foi organizado em quatro worksheets SQL distintas. Cada uma dessas etapas desempenha um papel fundamental na transformação e preparação dos dados para análise. A seguir, será detalhado o funcionamento de cada uma dessas zonas

### 4.1 Landing Zone - Staging Data

A Landing Zone é a etapa inicial do pipeline de dados, onde os dados são armazenados exatamente como foram extraídos da fonte original, sem qualquer tipo de tratamento ou transformação. Essa abordagem garante a preservação da integridade dos dados brutos, permitindo maior flexibilidade para diferentes formas de processamento e análise no futuro. Além disso, ao manter uma cópia fiel dos dados originais, essa zona possibilita auditorias, comparações históricas e correções em eventuais transformações realizadas nas fases posteriores do pipeline. Seguem os principais procedimentos realizados na Landing Zone:

- a) **Criação do Data Warehouse:** Nesta etapa, é realizada a criação do data warehouse, que atua como o cluster responsável pelo processamento dos dados. Para este projeto, o tamanho foi definido como 'X-Small', uma vez que não há necessidade de realizar tratamentos complexos nesta fase. Essa escolha reduz a capacidade de processamento, mas, por outro lado, minimiza os custos operacionais, o que seria um fator relevante em uma aplicação real.

```
1 create warehouse nba_staging_wh with warehouse_size='x-small';  
2
```

- b) **Criação do Banco de Dados:** Em seguida é realizada a criação banco de dados, que será utilizado para o armazenamento dos dados na Landing Zone.

```
3 create database if not exists nba_landing_db;  
4
```

- c) **Criação da Role:** Também é realizada a criação de uma role (função de acesso) específica para este projeto, a fim de gerenciar as permissões de uso dos dados. Essa role define quais usuários ou processos têm

autorização para acessar, inserir e manipular os dados, garantindo a segurança e o controle adequado dentro do ambiente do Snowflake assegurando a governança dos dados ao longo do pipeline.

```
5      create role if not exists nba_role;  
6
```

- d) **Criação do esquema:** Etapa onde é realizada a criação do esquema que irá receber as tabelas

```
7      create schema if not exists nba_landing_db.nba_landing_schema;  
8
```

- e) **Criação do stage:** Nesse procedimento, é realizada a criação do staging directory, que serve como um repositório temporário para armazenar os arquivos CSV antes que sejam processados e carregados nas tabelas do banco de dados. Esse diretório facilita a ingestão dos dados, permitindo que o Snowflake acesse os arquivos e os transforme em registros estruturados para posterior análise.

```
9      create stage nba_stage directory = ( enable = true ) encryption =  
10      ( type = 'SNOWFLAKE_SSE' );
```

- f) **Criação do formato de arquivo CSV:** Por padrão, é necessário definir o formato do arquivo a ser utilizado no Snowflake. Neste caso, foi criada uma configuração específica para arquivos no formato CSV, garantindo que os dados sejam lidos corretamente no stage e preservando aspectos como delimitadores e codificação de caracteres.

```
11      create or replace file format CSV  
12      type = csv  
13      field_delimiter = ','  
14      field_optionally_enclosed_by='"' --Evita considerar virgula  
dentro de strings  
15      skip_header = 1;  
16
```

- g) **Criação da estrutura das tabelas:** Nesta etapa, são criadas as tabelas que irão armazenar os dados extraídos dos arquivos CSV, garantindo que todos os campos sejam corretamente estruturados para posterior armazenamento e processamento. Abaixo, segue um exemplo de criação de tabela.

```
17      create or replace table player
18      (id integer          comment 'id Jogador',
19       full_name string     comment 'Nome completo',
20       first_name string    comment 'Nome',
21       last_name string     comment 'Sobrenome',
22       is_active integer   comment 'Em atividade'
23      );
24
```

- h) **Carregamento dos dados nas tabelas:** Por fim, os dados dos arquivos CSV, armazenados no stage, são carregados nas tabelas criadas na etapa anterior, garantindo que as informações sejam estruturadas e prontas para as próximas fases do pipeline de dados. Abaixo, segue um exemplo de carregamento de dados na tabela

```
25      copy into player from @nba_stage
26      files = ( 'player.csv' )
27      file_format = CSV;
28
```

## 4.2 Curated Zone - Tables

A Curated Zone é a etapa do pipeline onde os dados passam por um processo de refinamento, no qual são definidos seus tipos, limites e tamanhos apropriados. Esse ajuste otimiza o consumo de recursos do data warehouse, reduzindo a necessidade de armazenamento e melhorando a eficiência das consultas. Ao garantir que cada campo tenha um tipo de dado adequado e um tamanho otimizado, essa fase contribui para um processamento mais ágil e um melhor desempenho das análises realizadas posteriormente. Seguem os principais procedimentos realizados na Curated Zone:

- a) **Criação do Banco de Dados e Esquema:** Nesta etapa, não foi necessária a criação de um novo data warehouse ou role, pois serão utilizados os mesmos recursos configurados na primeira etapa. Foi criado apenas um novo banco de dados e schema para armazenar os dados refinados na Curated Zone.



```
29 create database if not exists nba_curated_db;  
30 create schema if not exists nba_curated_db.nba_curated_schema;  
31
```

- b) **Criação das tabelas curadas:** Nesta etapa, foram criadas as tabelas com tipos e tamanhos de dados bem definidos, baseadas nas tabelas da Landing Zone. Abaixo, segue um exemplo de criação de tabela

```
32 create or replace table player  
33 (id number(10,0) comment 'id Jogador',  
34 full_name varchar(40) comment 'Nome completo',  
35 first_name varchar(20) comment 'Nome',  
36 last_name varchar(30) comment 'Sobrenome',  
37 is_active number(1,0) comment 'Em atividade'  
38 );  
39
```

- c) **Carregamento dos dados nas tabelas curadas:** Por fim, os dados das tabelas da Landing Zone são carregados nas tabelas refinadas da Curated Zone, garantindo que as informações estejam estruturadas e otimizadas para análise e processamento posterior. Abaixo, segue um exemplo de carregamento de dados na tabela

```
40 insert into player  
41 select * from nba_landing_db.nba_landing_schema.player;  
42
```

### 4.3 Curated Zone - Data Processing

Especificamente para este projeto, foi realizada uma etapa de processamento de dados devido à presença de informações incompletas e à utilização de dois datasets provenientes de autores diferentes. Como resultado, foi necessário realizar ajustes manuais em certos campos, os quais serão detalhados a seguir.

- a) **Ajuste da tabela season\_stats:** Nesta etapa, diversos ajustes foram feitos na tabela season\_stats, que pertence ao dataset NBA Players Stats Since 1950, a fim de garantir a correta integração e vinculação com as tabelas do outro dataset.

O primeiro ajuste necessário foi com os nomes distintos entre as tabelas `season_stats` e `player`, pois o nome será utilizado para a geração das chaves entre as duas tabelas.

Exemplo de discrepância entre tabelas:

- `season_stats`: Em Bryant;
- `player`: Emmette Bryant.

```
43 update season_stats set player_name = 'Emmette Bryant' where  
44 player_name like '%Em Bryant%';
```

Após o ajuste dos nomes, foram definidos os ids dos jogadores da `season_stats` para realização da ligação com a tabela `player`.

```
45 update season_stats sst  
46 set sst.id_person = pl.id from  
47 (select pl.full_name, listagg(pl.id, ',') id from player pl  
group by (pl.full_name)) pl  
48 where pl.full_name = sst.player_name;  
49
```

Foi identificado que alguns jogadores estavam registrados como indivíduos distintos, mas com IDs diferentes, embora se tratassem da mesma pessoa. Para resolver isso, foi necessário realizar uma atualização manual em cada um desses casos, consolidando corretamente os registros e garantindo a integridade dos dados.

Exemplo de ajuste:

```
50 update season_stats set id_person = '77268' where idx in  
51 (548,686,833,962,1074,1288);
```

Para garantir o correto relacionamento com a tabela `team`, foi identificado que alguns times estavam com nomes abreviados inconsistentes. Esses nomes foram ajustados para garantir uniformidade e integridade nas ligações entre as tabelas.

Exemplo de ajuste:

```
52 update season_stats set team = 'PHX' where team = 'PHO';  
53
```

- b) **Ajuste da tabela team:** Também foi necessário o ajuste da tabela team pois foi verificado que nessa tabela existiam apenas os 30 times em atividade da NBA e não possuía o histórico, com isso, foi realizada a inserção manual dos times com os identificadores corretos para bater tanto com a tabela season\_stats quanto com a tabela game.

Exemplo de inserção:

```
54 insert into team Values  
55 (1610612767, 'Charlotte Bobcats', 'CHA', 'Bobcats', 'Charlotte',  
56 'North Carolina', 2004);
```

Ajuste de identificadores com a tabela game

```
57 update team tm  
58 set tm.id = gm.team_id_home from  
59 (select distinct gm.team_name_home, gm.team_id_home from game  
gm where gm.season_type not in ('Pre Season')) gm  
60 where tm.full_name = gm.team_name_home;  
61
```

## 4.4 Consumption Zone - Staging Data

Na etapa Consumption Zone, são criadas as tabelas de dimensões e fatos, que incluem as principais transformações realizadas nos dados. Essas tabelas são estruturadas de forma otimizada para garantir um consumo eficiente e eficaz no painel BI. As transformações realizadas visam garantir que os dados estejam prontos para análise e visualização, proporcionando uma experiência interativa e insights valiosos na plataforma de BI.

- a) **Criação do Data Warehouse Banco de Dados e Esquema:** Nesta etapa, foi criado um novo Data Warehouse com capacidade um pouco maior 'medium', visando melhorar a velocidade e performance na transformação dos dados no Snowflake, assim como otimizar a leitura e análise dos dados no Qlik. Além disso, foram configurados o banco de dados e o

schema específicos para suportar as operações na Consumption Zone, garantindo uma estrutura eficiente para o consumo de dados.

```
62 create or replace warehouse nba_analytics_wh with warehouse_size=  
    'medium';  
63 create database if not exists nba_consumption_db;  
64 create schema if not exists nba_consumption_db.  
    nba_consumption_schema;  
65
```

- b) **Criação das tabelas para consumo:** Nesta etapa, foram criadas as tabelas fato e dimensão que serão consumidas posteriormente.

Exemplo de criação:

```
66 create or replace table dim_time  
67 (  
68     pk_time number(12,0) comment 'id_time',  
69     nome_time varchar(40) comment 'Nome do time',  
70     nome_abreviado varchar(30) comment 'Nome abreviado do time',  
71     apelido varchar(30) comment 'Apelido do time',  
72     estado varchar(40) comment 'Estado do time',  
73     cidade varchar(40) comment 'Cidade do time',  
74     ano_fundacao number(4,0) comment 'Ano de Fundação do time'  
75 );  
76
```

- c) **Carregamento e transformação dos dados nas tabelas curadas:**  
Por fim, é realizado o carregamento e a transformação dos dados para alimentar as tabelas. Entre as principais transformações executadas, destacam-se:

- **Dim\_Jogador:** Ajuste de datas, valores nulos e unidades de medidas

```
77 insert into dim_jogador
78 select
79     Nvl(pl.is_active, 0) as em_atividade,
80     to_date(pli.birthdate) as data_nascimento,
81     Round(((Substr(height, 1, charindex('-',height)-1)*12) +
82     Substr(height, charindex('-',height)+1, Length(height))) as
83     altura_polegadas,
84     Round((((Substr(height, 1, charindex('-',height)-1)*12) +
85     Substr(height, charindex('-',height)+1, Length(height)))
86     *2.54) as altura_cm,
87     Round(pli.weight/2.205) as peso_kg,
88     ...
89 from nba_curated_db.nba_curated_schema.player pl
90 inner join nba_curated_db.nba_curated_schema.
common_player_info pli
on pl.id = pli.person_id
inner join nba_curated_db.nba_curated_schema.draft_history
dh
on pl.id = dh.person_id;
```

- **Dim\_Time:** Sem grandes transformações;
- **Dim\_Arbitro\_Jogo:** Sem grandes transformações;
- **Fato Stats Temporada:** No dataset original, a temporada da NBA é definida pelo ano de término. No entanto, como cada temporada se inicia em um ano e termina no seguinte, o formato foi ajustado para refletir ambos os anos, facilitando a análise.

```
91 insert into fato_stats_temporada
92 select
93     season_year-1 || '-' || Substr(season_year, 3, 4) as
94     temporada,
95     ...
96 from nba_curated_db.nba_curated_schema.season_stats sst
97 inner join nba_curated_db.nba_curated_schema.player pl
98     on pl.id = sst.id_person
99 inner join nba_curated_db.nba_curated_schema.team tm
100     on sst.team = tm.abbreviation
101 order by sst.season_year asc;
```

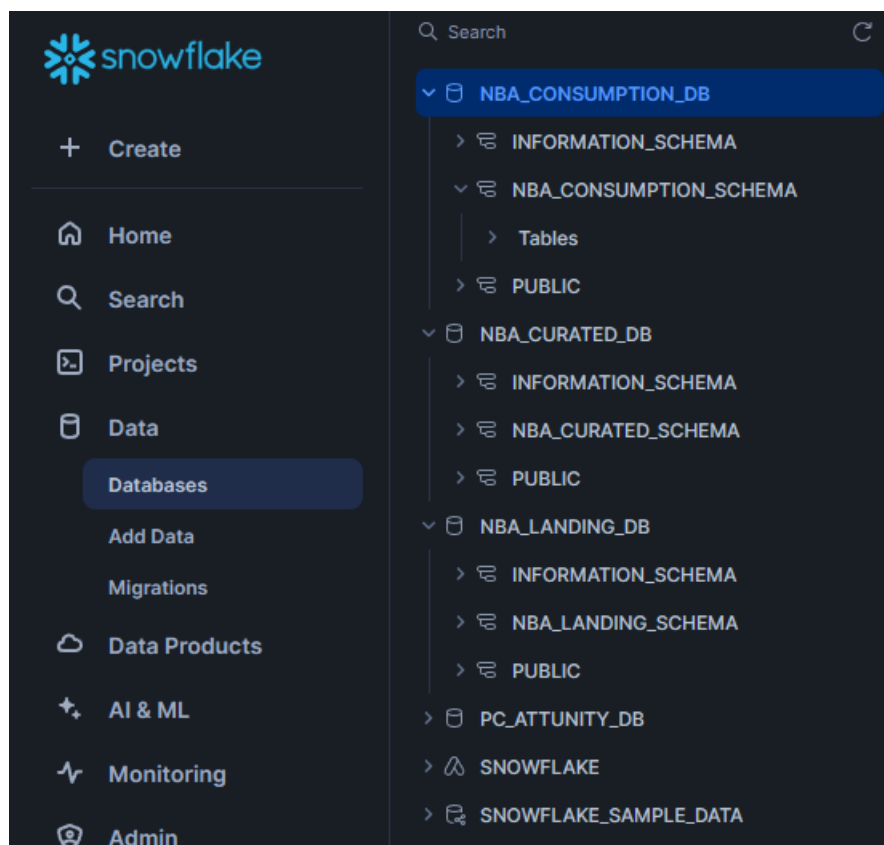
- **Fato\_Jogo:** Além do ajuste relacionado à temporada, a transformação mais significativa foi a reestruturação da tabela para dividir os times que participaram de um jogo em múltiplos registros. No formato original, cada registro continha informações de ambos os times nos campos `team_home` e `team_away`. Para facilitar as análises, a estrutura foi modificada da seguinte forma:
  - Foram criadas duas tabelas temporárias com a mesma estrutura: uma contendo os times que jogaram em casa (home) e outra com os times visitantes (away).
  - Um novo campo foi adicionado para indicar a posição do time no jogo (Casa ou Visitante), possibilitando a aplicação de filtros e permitindo uma análise mais detalhada.

```
102 insert into fato_jogo
103 --Time Casa
104 select
105     gm.game_id          as pk_jogo,
106     gm.team_id_away     as pk_time,
107     gms.season || '-' || Substr(gms.season+1, 3, 4) as temporada
108     ...
109 from nba_curated_db.nba_curated_schema.game gm
110 left join nba_curated_db.nba_curated_schema.game_summary gms
111     on gm.game_id = gms.game_id
112 where gm.season_type not in ('Pre Season', 'All Star', 'All-Star
113                                ')
114 UNION ALL
115 --Time Visitante
116 select
117     gm.game_id          as pk_jogo,
118     gm.team_id_home     as pk_time,
119     gms.season || '-' || Substr(gms.season+1, 3, 4) as temporada
120     ...
121 from nba_curated_db.nba_curated_schema.game gm
122 left join nba_curated_db.nba_curated_schema.game_summary gms
123     on gm.game_id = gms.game_id
124 where gm.season_type not in ('Pre Season', 'All Star', 'All-Star
                                ');
```

Com a Consumption Zone finalizada, os dados estão prontos para serem consumidos em qualquer painel de análise. Neste projeto, optou-se pelo Qlik como a plataforma para visualização e exploração dos dados. O Qlik oferece a flexibilidade

necessária para criar dashboards interativos e gerar insights a partir das informações tratadas.

**Figura 2 – Estrutura Final Banco de Dados - Snowflake**



Para um detalhamento completo dos scripts utilizados durante o pipeline, os mesmos estão disponíveis no GitHub do projeto, onde é possível acessar o código-fonte e acompanhar a implementação de cada etapa.

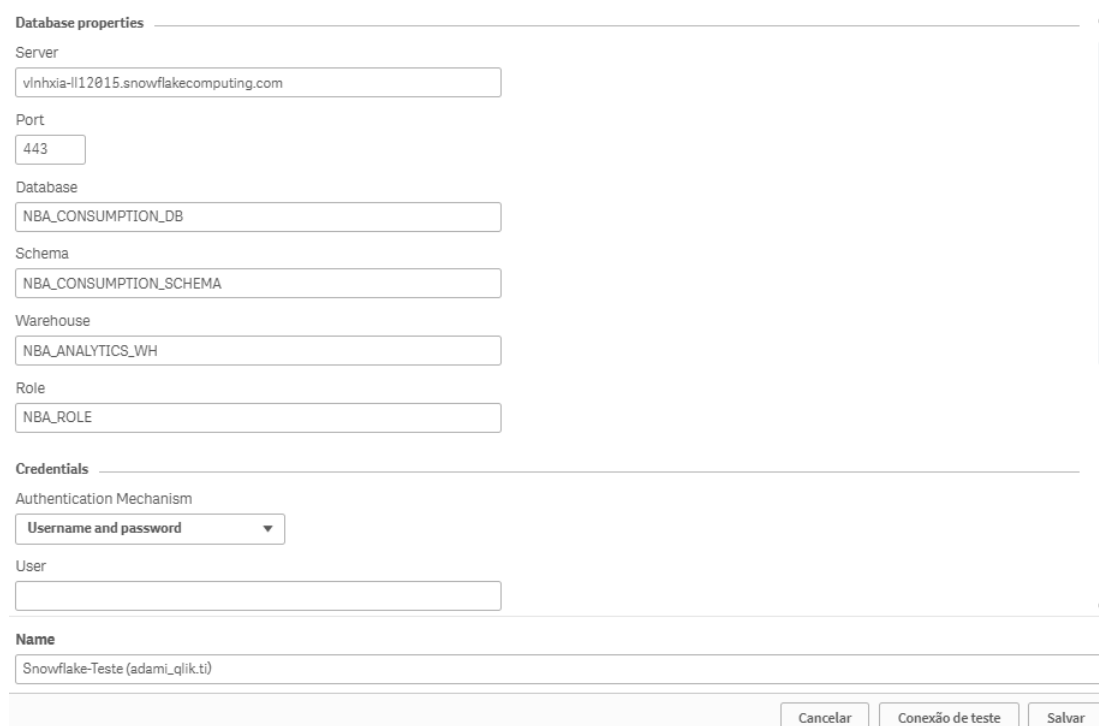
## 5 ETL EXTRA E PAINÉIS - QLIK CLOUD

Para o consumo de dados na plataforma Qlik, é necessário um processo de ETL extra e com isso é possível sintetizar ainda mais os dados e focar nos insights que serão desenvolvidos dentro da plataforma.

### 5.1 Conexão com o Snowflake

Primeiramente, é necessário estabelecer a conexão entre o Qlik e o Snowflake, conectando-se ao data warehouse previamente configurado. Isso permite que os dados da Consumption Zone sejam acessados de forma eficiente. Essa conexão é fundamental para garantir que os dados sejam extraídos corretamente e preparados para análise dentro do Qlik. O processo é ilustrado na Figura 3.

**Figura 3 – Conexão de dados Snowflake > Qlik**



The screenshot displays the 'Database properties' and 'Credentials' configuration windows in Qlik Cloud. The 'Database properties' section includes fields for Server (vinhxia-ll12015.snowflakecomputing.com), Port (443), Database (NBA\_CONSUMPTION\_DB), Schema (NBA\_CONSUMPTION\_SCHEMA), Warehouse (NBA\_ANALYTICS\_WH), and Role (NBA\_ROLE). The 'Credentials' section shows the Authentication Mechanism set to 'Username and password', with empty fields for User and Password. A 'Name' field at the bottom is populated with 'Snowflake-Teste (adami\_qlik.ti)'. At the bottom right, there are three buttons: 'Cancelar', 'Conexão de teste', and 'Salvar'.

### 5.2 Carregando dados

#### 5.2.1 Extração

Uma vez estabelecida a conexão, os dados foram extraídos do Snowflake e armazenados no Qlik no formato .qvd. Este formato é altamente eficiente para armazenar grandes volumes de dados, garantindo rapidez no acesso e na carga dos dados em futuras análises. O uso do .qvd facilita a manipulação e a performance



durante a criação de visualizações e dashboards no Qlik, além de ser ideal para o consumo de dados em tempo real ou análises históricas.

```
125 //Conexão snowflake
126 LIB CONNECT TO 'SnowFlake WH';
127 //
128 [DIM_JOGADOR]:
129 SELECT * FROM "NBA_CONSUMPTION_DB"."NBA_CONSUMPTION_SCHEMA"."DIM_JOGADOR";
130 Store [DIM_JOGADOR] into [lib://DataFiles/Extraídos/NBA/DIM_JOGADOR.qvd] (qvd);
131 Drop Table [DIM_JOGADOR];
132
133 //
134 [DIM_TIME]:
135 SELECT * FROM "NBA_CONSUMPTION_DB"."NBA_CONSUMPTION_SCHEMA"."DIM_TIME";
136 Store [DIM_TIME] into [lib://DataFiles/Extraídos/NBA/DIM_TIME.qvd] (qvd);
137 Drop Table [DIM_TIME];
138
139 //
140 [DIM_ARBITRO_JOGO]:
141 SELECT * FROM "NBA_CONSUMPTION_DB"."NBA_CONSUMPTION_SCHEMA"."DIM_ARBITRO_JOGO";
142 Store [DIM_ARBITRO_JOGO] into [lib://DataFiles/Extraídos/NBA/DIM_ARBITRO_JOGO.qvd]
    (qvd);
143 Drop Table [DIM_ARBITRO_JOGO];
144
145 //
146 [FATO_JOGO]:
147 SELECT * FROM "NBA_CONSUMPTION_DB"."NBA_CONSUMPTION_SCHEMA"."FATO_JOGO";
148 Store [FATO_JOGO] into [lib://DataFiles/Extraídos/NBA/FATO_JOGO.qvd] (qvd);
149 Drop Table [FATO_JOGO];
150
151 //
152 [FATO_STATS_TEMPORADA]:
153 SELECT * FROM "NBA_CONSUMPTION_DB"."NBA_CONSUMPTION_SCHEMA"."FATO_STATS_TEMPORADA";
154 Store [FATO_STATS_TEMPORADA] into [lib://DataFiles/Extraídos/NBA/
    FATO_STATS_TEMPORADA.qvd] (qvd);
155 Drop Table [FATO_STATS_TEMPORADA];
156
```

### 5.2.2 Transformação

Após a extração dos dados, foram realizadas transformações adicionais diretamente no sistema Qlik, com foco na obtenção de insights específicos para o projeto. Como o dataset NBA Players Stats Since 1950 abrange dados apenas até a temporada 2016-17, foi necessário limitar os dados de todas as tabelas até a temporada

de 2017. Isso garantiu que as análises estivessem alinhadas com o período disponível no dataset.

Exemplo:

```
157 //
158 Jogos_tmp:
159 LOAD
160     PK_JOGO          as PK_Jogo,
161     PK_TIME          as PK_Time,
162     POSICAO_TIME     as [Jogo - Posição Time],
163     If(IsNull(TEMPORADA), If(Month(DATA_JOGO) >= 10,
164                             Year(DATA_JOGO) & '-' & Mid((Year(DATA_JOGO) + 1), 3),
165                             (Year(DATA_JOGO) - 1) & '-' & Mid(Year(DATA_JOGO), 3)),
166     TEMPORADA) as [Jogo - Temporada],
167     ...
168 FROM [lib://DataFiles/Extraidos/NBA/FATO_JOGO.qvd] (qvd)
169 Where Date(DATA_JOGO, 'DD/MM/YYYY') <= '12/06/2017';
170
```

Além disso, foram criadas novas dimensões para enriquecer as visualizações e facilitar a análise:

1. **Dimensão de Times Campeões:** Esta dimensão foi criada para destacar as equipes campeãs da NBA, permitindo a geração de insights sobre o histórico de vitórias e o desempenho de times vencedores ao longo do tempo, a definição do campeão foi realizada com base no vencedor do último jogo dos playoffs da temporada.

```
171 Times_Campeões:
172 Load
173     "Jogo - Temporada",
174     Date(Max("Jogo - Data"), 'DD/MM/YYYY') as [Campeão - Último Jogo
        Temporada]
175 From [lib://DataFiles/Transformados/NBA/Jogos.qvd] (qvd)
176 Where
177     "Jogo - Tipo Temporada" = 'Playoffs'
178     and "Jogo - Ganhou/Perdeu" = 'W'
179 Group By "Jogo - Temporada";
180 Left Join(Times_Campeões)
181 Load
182     "Jogo - Data" as [Campeão - Último Jogo Temporada],
183     PK_Time
184 From [lib://DataFiles/Transformados/NBA/Jogos.qvd] (qvd)
185 Where
186     "Jogo - Tipo Temporada" = 'Playoffs'
187     and "Jogo - Ganhou/Perdeu" = 'W';
188
189 //Campeões não registrados
190 Concatenate (Times_Campeões)
191 Load * Inline
192 [
193     PK_Time, Campeão - Último Jogo Temporada
194     1610612786, 12/04/1958
195     1610612738, 09/04/1959
196     1610612738, 25/04/1965
197     1610612738, 18/04/1961
198     1610612738, 05/05/1969
199     1610612745, 22/06/1994
200     1610612741, 16/06/1996
201     1610612747, 19/06/2000
202     1610612747, 12/06/2002
203     1610612748, 20/06/2006
204 ];
205
```

2. **Dimensão de Estatísticas dos Jogadores (radar):** Foi criada uma dimensão específica para as estatísticas dos jogadores, que possibilitou a construção de um radar. Possibilitando a comparação de desempenho entre jogadores.

```
206 //
207 Estatísticas_Temporada_Jogador_Radar:
208 Load
209     PK_Jogador &'| '& Temporada as PK_Jogador_Temporada,
210     'Arremesso 3pt' as [EstatísticasRadar - Categoria],
211     ([Estatística - Acertos 3pt] - [EstatísticaMin - Acertos 3pt]) /
212     ([EstatísticaMax - Acertos 3pt] - [EstatísticaMin - Acertos 3pt]) as [
213     EstatísticasRadar - Valor]
213 From [lib://DataFiles/Transformados/NBA/Estatísticas_Temporada_Jogador.qvd] (
214     qvd)
214 Where [Estatística - Acertos 3pt] > 0;
215
216 Concatenate(Estatísticas_Temporada_Jogador_Radar)
217
218 Load
219     PK_Jogador &'| '& Temporada as PK_Jogador_Temporada,
220     'Arremesso 2pt' as [EstatísticasRadar - Categoria
221     ],
222     ([Estatística - Acertos 2pt] - [EstatísticaMin - Acertos 2pt]) /
223     ([EstatísticaMax - Acertos 2pt] - [EstatísticaMin - Acertos 2pt]) as [
224     EstatísticasRadar - Valor]
223 From [lib://DataFiles/Transformados/NBA/Estatísticas_Temporada_Jogador.qvd] (
224     qvd)
224 Where [Estatística - Acertos 2pt] > 0;
225 ...
226
```

### 5.2.3 Carregamento

Finalmente, os dados transformados são carregados no Qlik, prontos para serem utilizados nas visualizações e painéis interativos. Esse processo de carregamento garante que todas as informações estejam estruturadas de maneira otimizada para análise.

```
227 //
228 Estatísticas_Temporada_Jogador:
229 LOAD * FROM [lib://DataFiles/Transformados/NBA/Estatísticas_Temporada_Jogador.qvd] (
      qvd);
230
231 Jogadores:
232 LOAD * FROM [lib://DataFiles/Transformados/NBA/Jogadores.qvd] (qvd)
233 Where Exists (PK_Jogador);
234
235 //
236 Estatísticas_Temporada_Jogador_Radar:
237 LOAD * FROM [lib://DataFiles/Transformados/NBA/Estatísticas_Temporada_Jogador_Radar
      .qvd] (qvd);
238
239 //
240 Times:
241 LOAD * FROM [lib://DataFiles/Transformados/NBA/Times.qvd] (qvd);
242
243 //
244 Times_Campeões:
245 LOAD * FROM [lib://DataFiles/Transformados/NBA/Times_Campeões.qvd] (qvd);
246
247 //
248 Jogos:
249 LOAD * FROM [lib://DataFiles/Transformados/NBA/Jogos.qvd] (qvd);
250
251 //
252 Arbitros_Jogo:
253 LOAD * FROM [lib://DataFiles/Transformados/NBA/Arbitros_Jogo.qvd] (qvd);
254
```

Figura 4 – Modelagem Final - Qlik

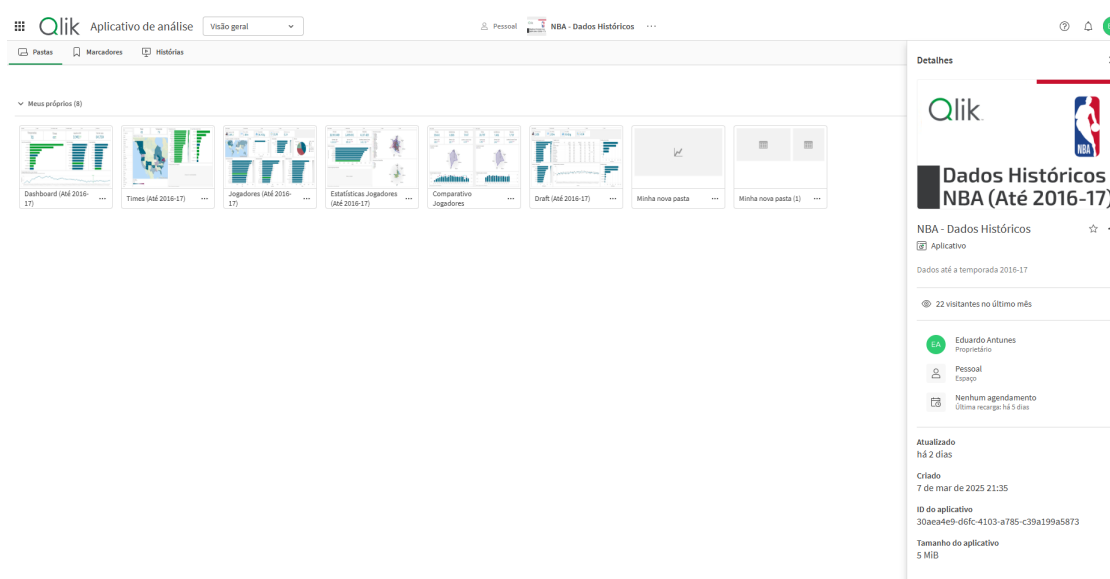


### 5.3 Criação do painel BI

Com os dados devidamente tratados e carregados no Qlik Cloud, foram desenvolvidos dashboards interativos para a exploração dos dados, possibilitando a obtenção de insights.

A seguir, são apresentadas as visões e telas desenvolvidas. Além disso, algumas dessas visualizações serão utilizadas para responder às questões levantadas no escopo do projeto, proporcionando uma análise mais aprofundada dos padrões e tendências identificados.

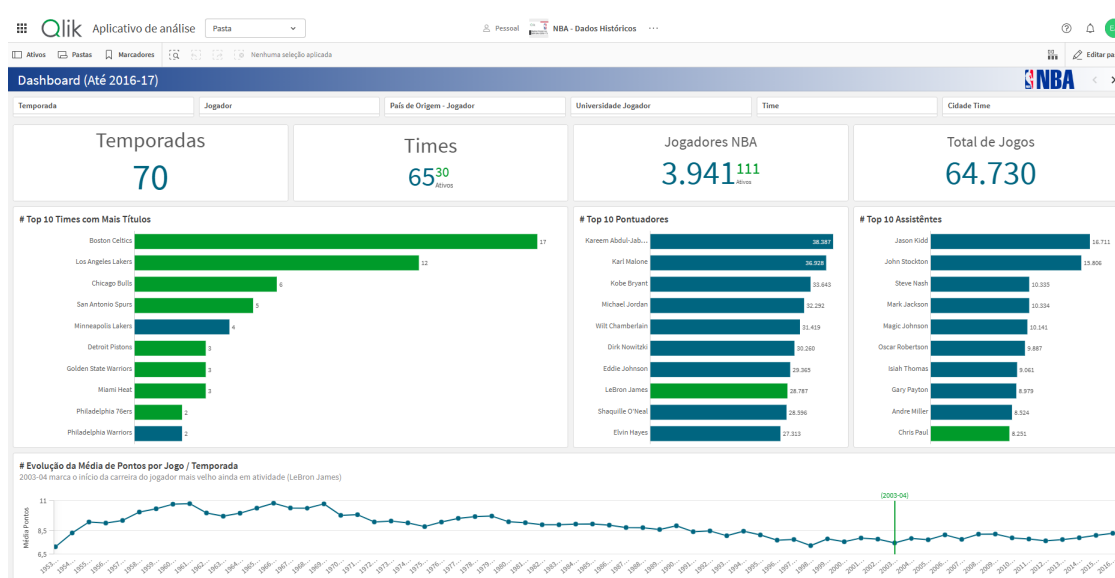
Figura 5 – Visão Geral Painel - NBA Dados Históricos



### 5.3.1 Dashboard

A tela inicial apresenta uma visão geral com dados essenciais sobre jogadores e times da NBA, funcionando como um ponto de partida para a exploração dos dados. Um conceito central adotado em toda a aplicação é o uso da cor verde para indicar jogadores e times que ainda estão ativos na NBA. Essa abordagem facilita a comparação rápida entre atletas em atividade e jogadores aposentados, assim como entre franquias em operação e equipes extintas. Além disso, essa tela oferece estatísticas resumidas e filtros interativos, permitindo uma navegação intuitiva e uma melhor contextualização das informações antes de explorar análises mais detalhadas nos painéis subsequentes.

**Figura 6 – Dashboard - NBA Dados Históricos**



### 5.3.2 Times

Esta tela apresenta os principais dados de todas as equipes que já passaram pela NBA, incluindo informações como nome, histórico de temporadas e estado de origem de cada time.

Entre as análises disponíveis, destaca-se o gráfico "Evolução dos Times por Temporada", que permite visualizar o desempenho e a trajetória de um time específico ao longo dos anos. Para uma interpretação mais clara, essa análise exige a seleção de um único time nos filtros, pois a comparação simultânea de múltiplas equipes comprometeria a clareza da visualização.

Figura 7 – Times Sem Filtro - NBA Dados Históricos

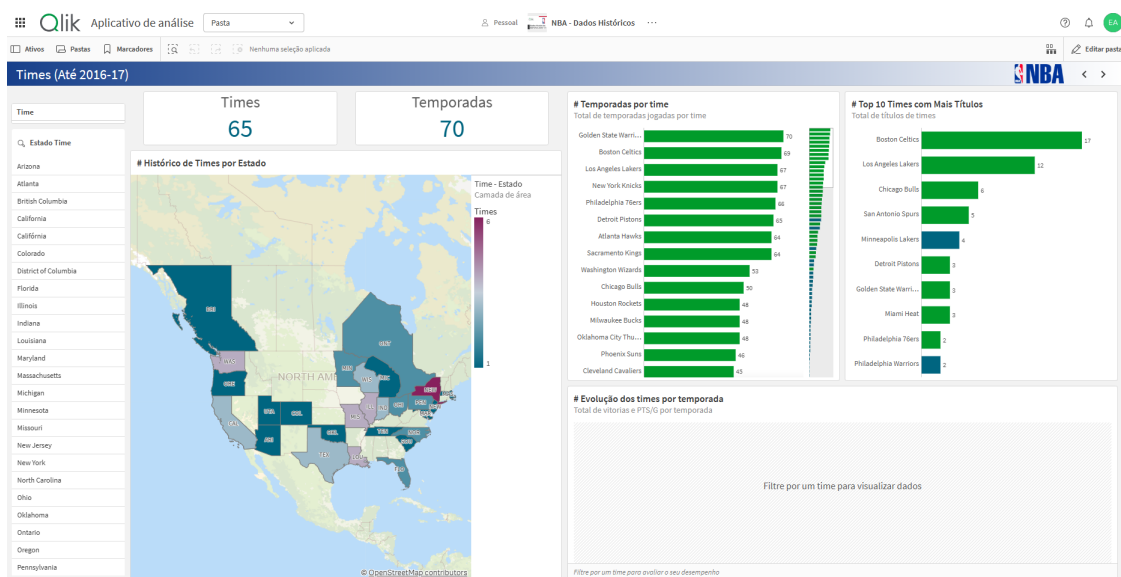


Figura 8 – Times Com Filtro - NBA Dados Históricos

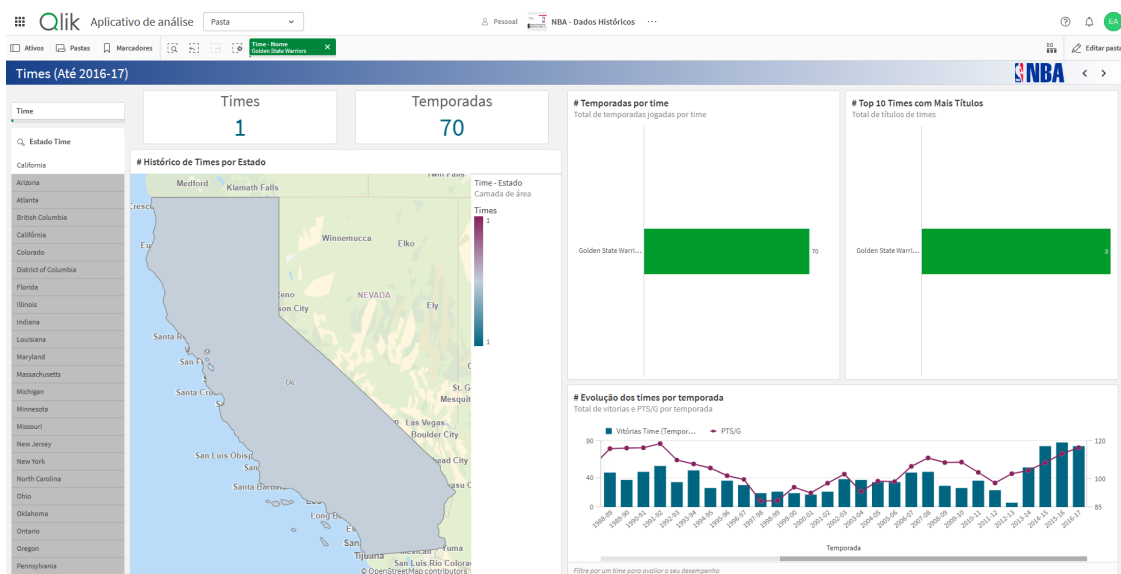
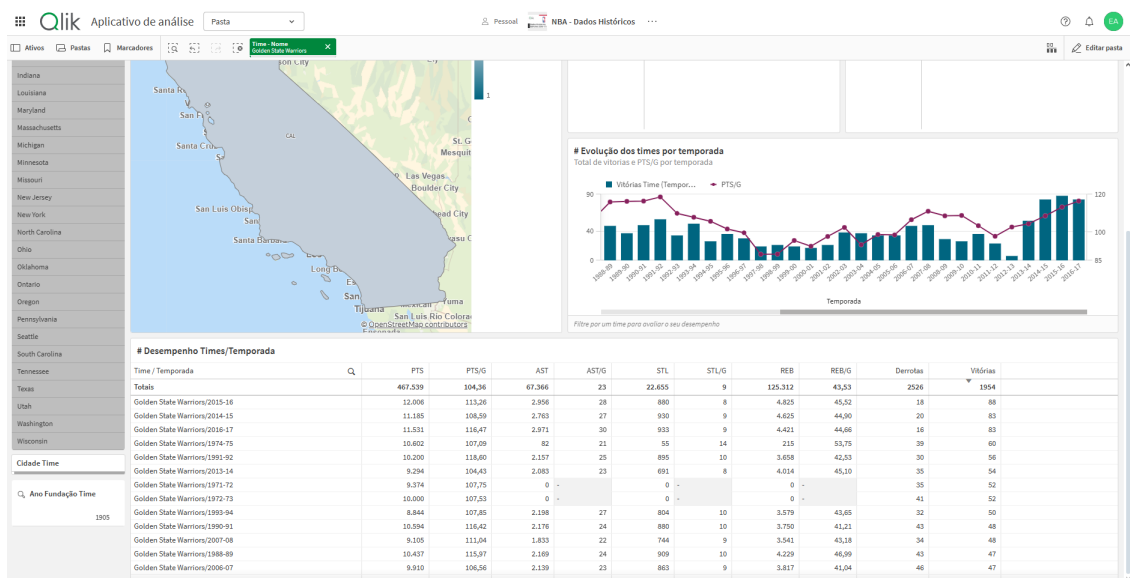




Figura 9 – Times Analítico - NBA Dados Históricos



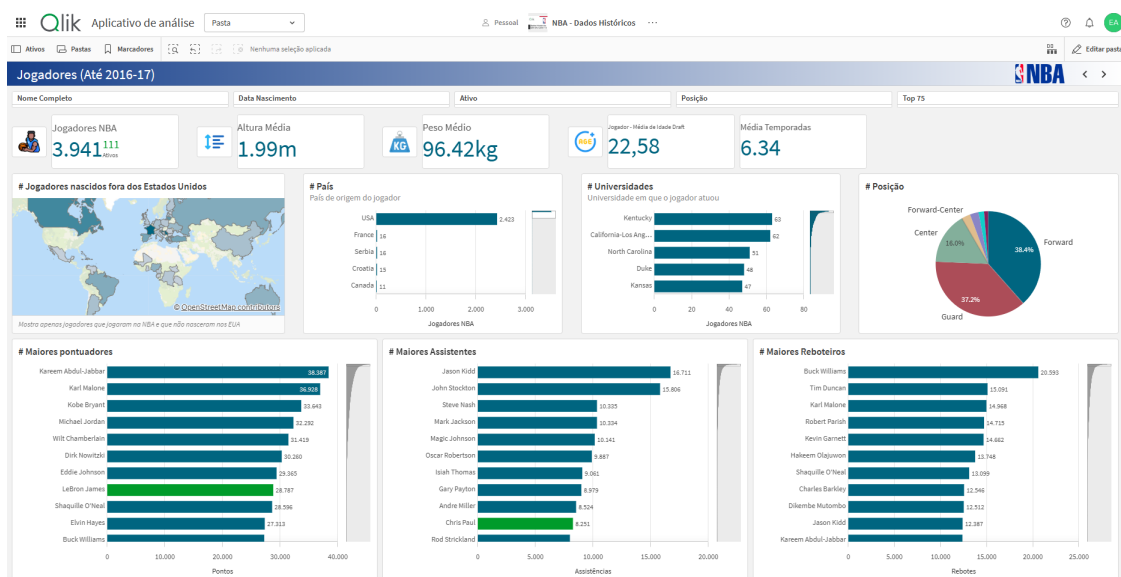
### 5.3.3 Jogadores

Nesta tela, foram elaborados diversos gráficos gerais para análise dos jogadores, proporcionando uma visão abrangente de sua distribuição e trajetória. Entre as visualizações disponíveis, destacam-se:

- Mapa geográfico mostrando a localidade de nascimento dos jogadores.
- Gráfico de universidades indicando as instituições que mais formaram atletas da NBA.
- Distribuição de posições para análise do perfil dos jogadores ao longo dos anos.
- Rankings estatísticos permitindo identificar os líderes em diferentes métricas, como pontos, assistências e rebotes.

Essas análises auxiliam na compreensão dos padrões históricos e das influências regionais na formação dos atletas.

Figura 10 – Jogadores - NBA Dados Históricos



### 5.3.4 Estatísticas Jogadores

Com os insights desenvolvidos nos dashboards, é possível realizar uma análise aprofundada do desempenho de cada jogador, aplicando filtros personalizados como temporada, equipe, posição e estatísticas individuais.

Essa funcionalidade permite identificar tendências ao longo dos anos, comparar atletas em diferentes períodos e equipes, além de correlacionar seu desempenho com fatores estratégicos, como tempo de jogo, posição em quadra.

Figura 11 – Estatísticas Sem Filtro - NBA Dados Históricos

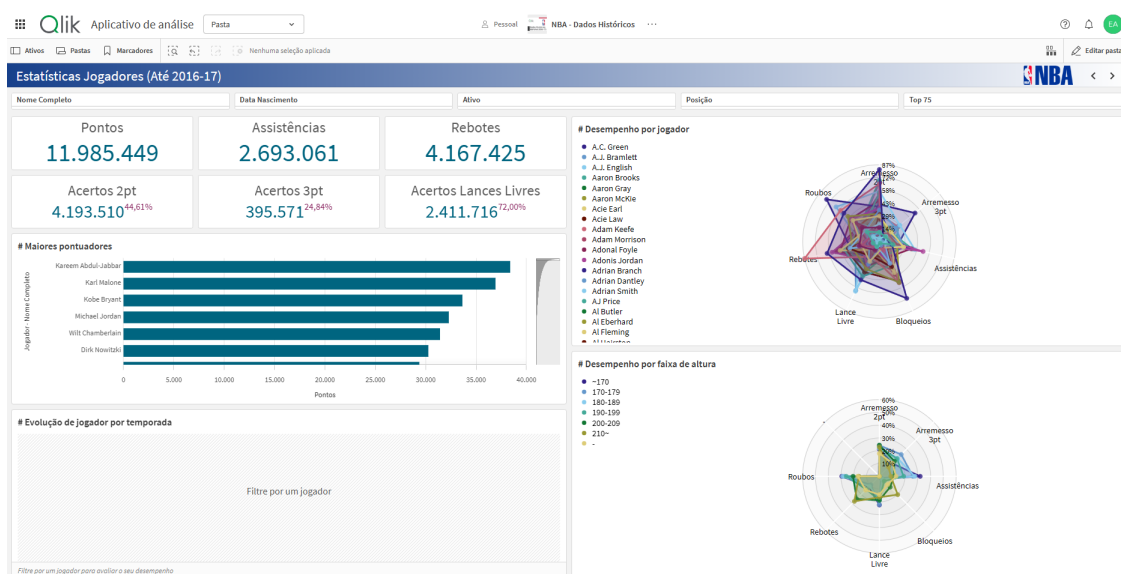
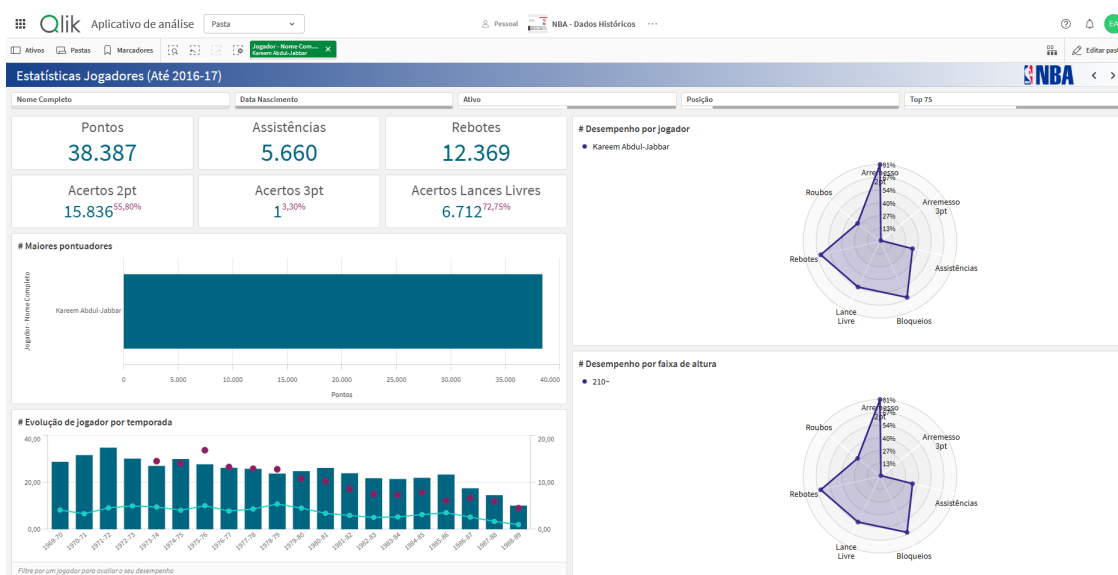


Figura 12 – Estatísticas Com Filtro - NBA Dados Históricos



### 5.3.5 Comparativo Jogadores

Foi desenvolvida uma tela interativa que permite a comparação detalhada entre dois ou mais jogadores, exibindo suas estatísticas individuais em diferentes temporadas. Essa funcionalidade possibilita a análise de desempenho com métricas como pontos, assistências, rebotes, entre outras.

Além disso, gráficos e tabelas dinâmicas foram implementados gráficos de radar para facilitar a visualização das diferenças e semelhanças entre os atletas, auxiliando na identificação de padrões e tendências ao longo de suas carreiras.

Figura 13 – Comparativo Sem Filtro - NBA Dados Históricos

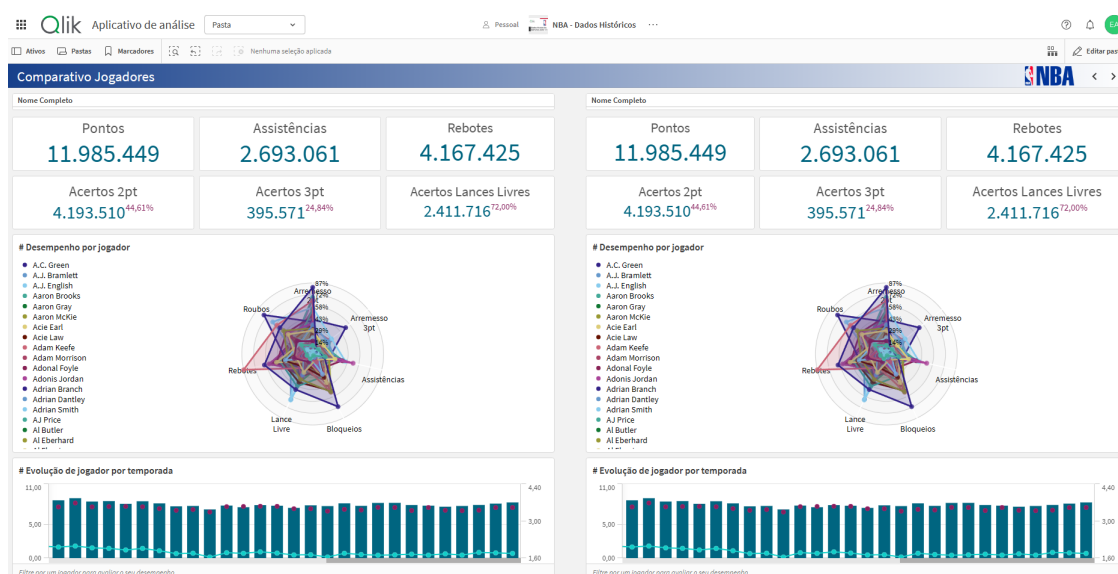
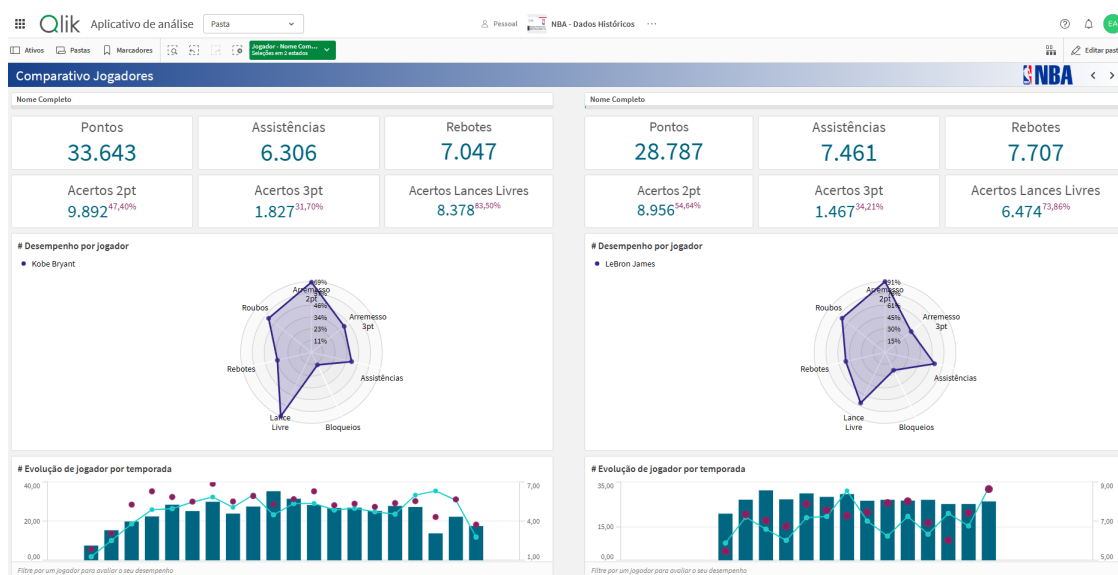


Figura 14 – Comparativo Com Filtro - NBA Dados Históricos

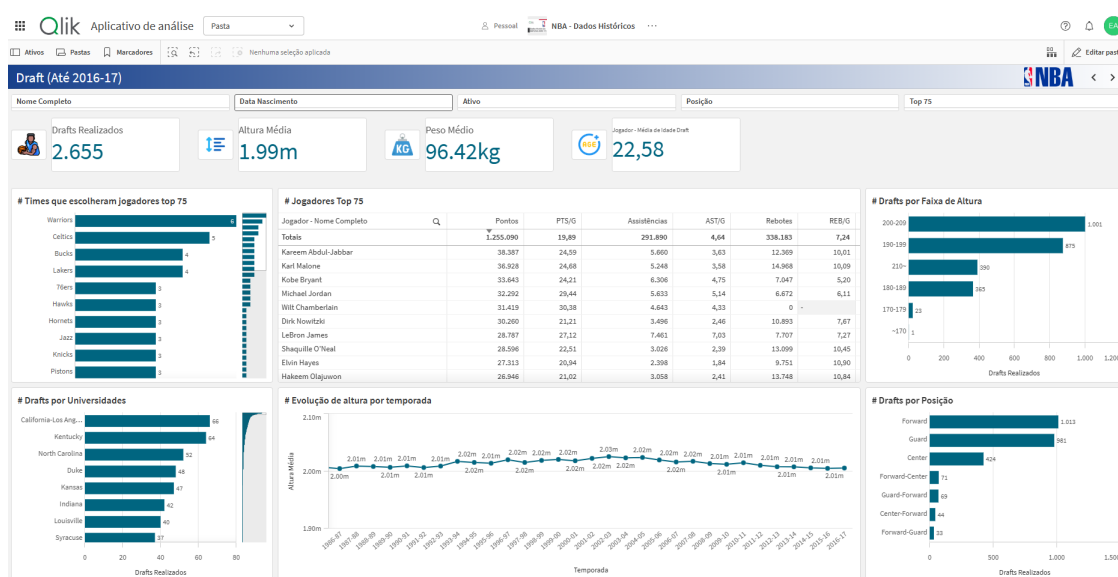


### 5.3.6 Draft

Por fim, foi desenvolvido um painel dedicado à análise dos dados do Draft da NBA, abrangendo informações detalhadas sobre todos os jogadores selecionados. Esse painel inclui métricas como altura, posição em quadra, universidades de origem, além de destacar jogadores que integram a lista dos Top 75 da NBA.

Com essa visualização, é possível identificar padrões e tendências ao longo dos anos, como a evolução do perfil dos atletas recrutados e a influência de determinadas universidades na formação de talentos da liga.

Figura 15 – Draft - NBA Dados Históricos



Com as telas desenvolvidas, foi possível criar uma solução interativa e abrangente para a análise dos dados históricos da NBA. O painel fornece visões detalhadas sobre jogadores, times, drafts e estatísticas, permitindo a exploração dos dados de forma dinâmica.

Cada tela foi estruturada para atender a diferentes necessidades analíticas, desde uma visão geral até comparações individuais entre jogadores e times. Destacam-se as funcionalidades que possibilitam identificar padrões nos drafts, avaliar o impacto de características físicas no desempenho dos atletas e explorar estatísticas consolidadas ao longo das temporadas.

6 ANÁLISES, RESULTADOS OBTIDOS E CONCLUSÕES

Através das telas desenvolvidas, pode-se obter respostas para as perguntas levantadas anteriormente no escopo do projeto. Com a interação nos dashboards, os usuários podem realizar diversas análises.

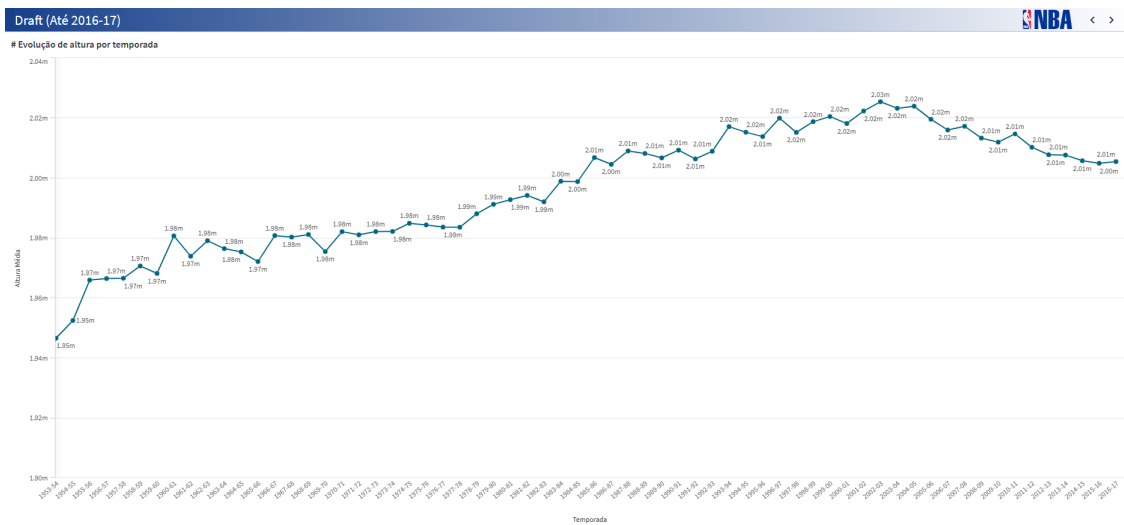
6.1 Respostas

A seguir, serão apresentadas as respostas para as questões-chave definidas no escopo do projeto, com base nas análises realizadas ao longo do estudo. Cada pergunta será explorada utilizando os dados, visualizações e insights extraídos do painel desenvolvido, permitindo uma compreensão mais aprofundada sobre os aspectos históricos e estatísticos da NBA.

6.1.1 Quais as tendências e padrões em Drafts da NBA?

Conforme a Figura 16 utilizando o painel Draft foi possível identificar uma tendência ao longo dos anos, onde a média de altura dos jogadores escolhidos nos drafts da NBA tem aumentado. Esse padrão sugere que as equipes passaram a priorizar jogadores mais altos, possivelmente devido à percepção de que atletas com maior estatura oferecem vantagens em aspectos cruciais do jogo, como o controle do rebote, a defesa e a presença no garrafão. Esse dado reflete mudanças nas estratégias de recrutamento e nas características valorizadas pelos times na formação de seus elencos. No painel de Draft (Figura 15) é possível avaliar diversas tendências relacionadas a universidades e posição de jogadores.

Figura 16 – Draft Altura - NBA Dados Históricos



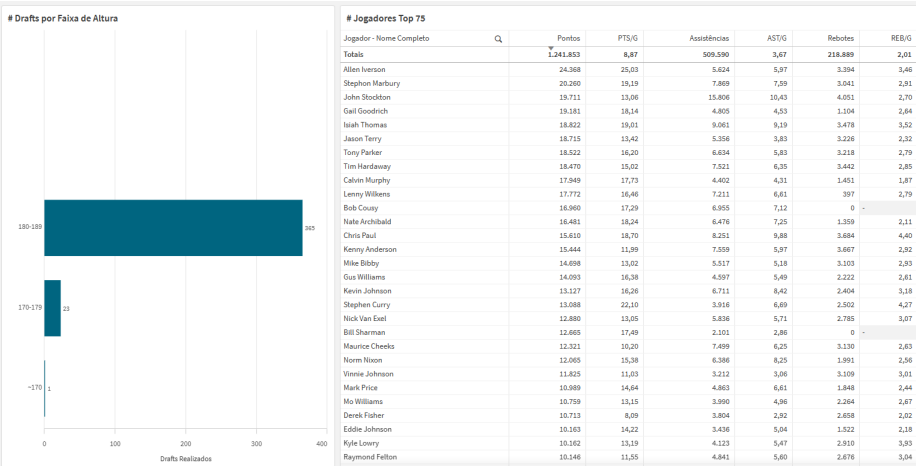
6.1.2 Existe um impacto da altura dos jogadores no desempenho em quadra?

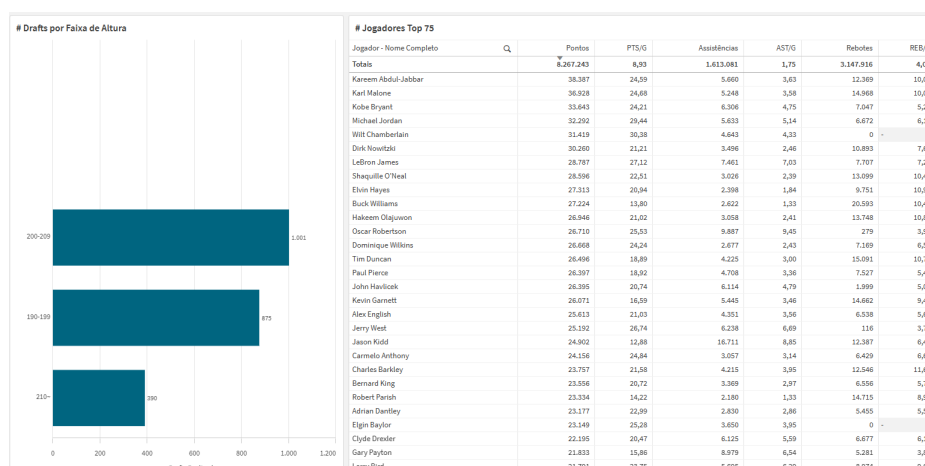
No contexto da NBA, é importante considerar que existe uma tendência clara das equipes priorizarem jogadores com altura superior a 190 cm. Essa preferência pode influenciar diretamente as análises de desempenho relacionadas à altura. No entanto, para uma avaliação mais objetiva desse impacto, é possível comparar as médias de pontos, assistências e rebotes por jogo entre dois grupos distintos: jogadores com altura inferior a 190 cm e jogadores com 190 cm ou mais.

Com base nas Figuras 17 e 18, observa-se que a média de pontos por jogo entre os dois grupos não apresenta uma diferença tão expressiva. Por outro lado, as médias de assistências e rebotes por jogo revelam diferenças mais significativas, com os jogadores mais altos apresentando desempenho superior, especialmente em rebotes.

Esse padrão ajuda a justificar a preferência por atletas mais altos no recrutamento e desenvolvimento de talentos. Na NBA moderna, espera-se que os jogadores sejam cada vez mais versáteis, com capacidade de contribuir tanto ofensivamente quanto defensivamente. Assim, características físicas como a altura acabam se tornando um diferencial estratégico no desempenho coletivo e individual dentro de quadra.

Figura 17 – Estatísticas de jogadores menores que 190 cm - NBA Dados Históricos



**Figura 18 – Estatísticas de jogadores com 190cm ou mais - NBA Dados Históricos**

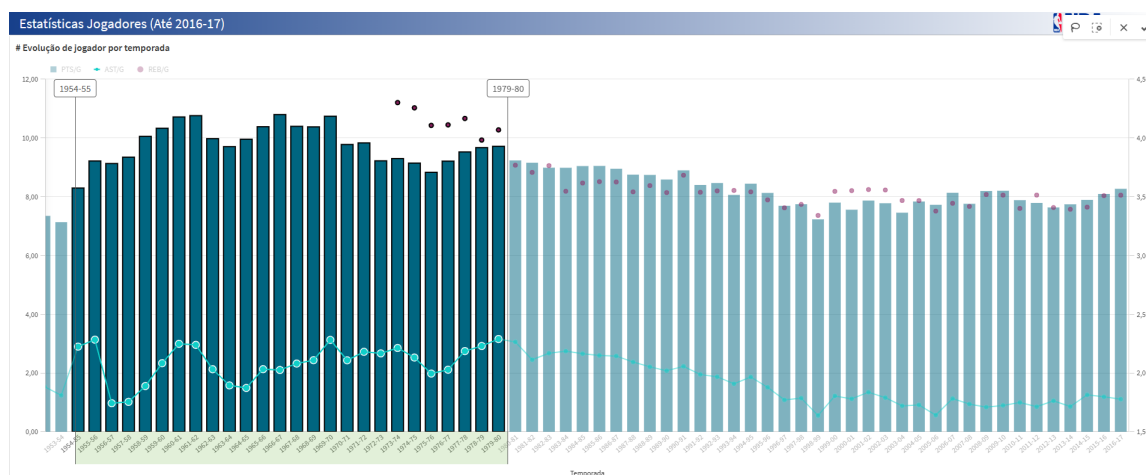
### 6.1.3 Como as estatísticas dos jogadores evoluíram ao longo do tempo?

A análise da média geral das estatísticas dos jogadores por temporada, considerando indicadores como PTS/G (Pontos por Jogo), AST/G (Assistências por Jogo) e REB/G (Rebotes por Jogo), permite identificar mudanças relevantes ao longo da história da NBA.

Os dados de rebotes por jogo (REB/G) estão disponíveis apenas a partir da temporada 1973-74, o que limita comparações anteriores. A partir da temporada 1980-81, observa-se uma queda significativa nas médias de PTS/G e REB/G, seguida de uma estabilização com menor variação ao longo dos anos seguintes. Essa mudança pode indicar que, antes da década de 1980, não era comum registrar estatísticas de forma tão padronizada para todos os jogadores, o que pode ter inflado as médias gerais, já que eram contabilizados apenas atletas com desempenhos mais destacados.

Na Figura 19, observa-se claramente esse comportamento, especialmente no intervalo entre as temporadas 1954-55 e 1979-80, período em que as médias apresentavam maiores variações e valores mais elevados em comparação às temporadas posteriores. Essa análise reforça a evolução na coleta e padronização dos dados estatísticos ao longo do tempo.



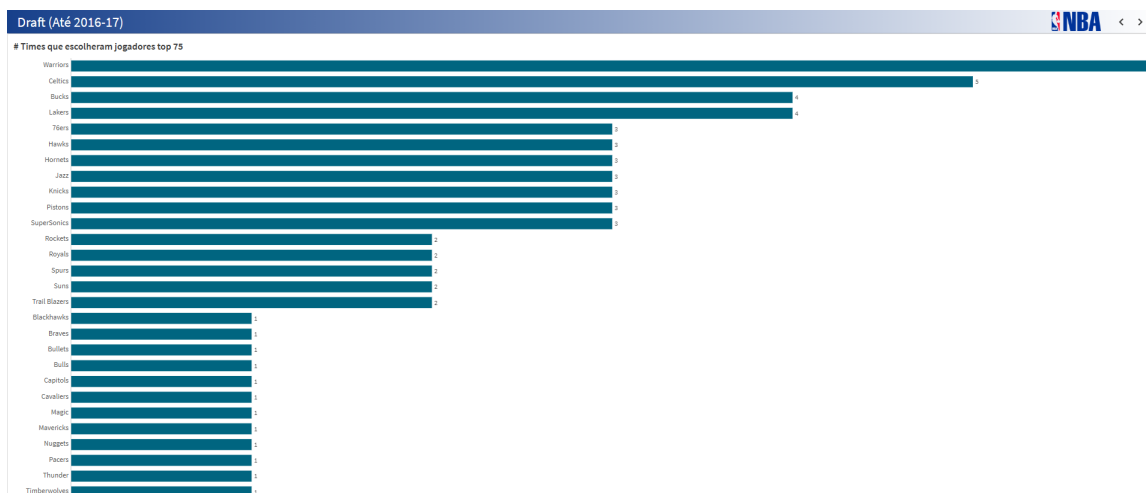
**Figura 19 – Evolução Estatísticas - NBA Dados Históricos**

#### 6.1.4 Quais equipes tiveram maior impacto na revelação de grandes estrelas?

A base de dados utilizada neste projeto conta com uma flag indicativa dos jogadores que integram a lista dos 75 maiores da história da NBA. A partir dessa informação, foi desenvolvido um gráfico (Figura 20) que mostra quais equipes selecionaram esses atletas no Draft ao longo dos anos.

A análise revela que o Golden State Warriors é a equipe que mais escolheu jogadores considerados grandes estrelas, com 6 atletas presentes na lista. Em seguida, aparecem o Boston Celtics, com 5 jogadores, e tanto o Milwaukee Bucks quanto o Los Angeles Lakers, com 4 jogadores cada.

No total, os 75 jogadores foram selecionados por 28 franquias diferentes, o que demonstra uma relativa dispersão desses talentos entre os times, mas também destaca algumas equipes como particularmente bem-sucedidas na identificação e seleção de grandes talentos no Draft.

**Figura 20 – Draft Times que Escolheram Grandes Estrelas - NBA Dados Históricos**

Essas são apenas algumas das perguntas que podem ser respondidas com base neste conjunto de dados. A riqueza e profundidade das informações disponíveis permitem inúmeras outras análises, que podem ser exploradas de forma mais detalhada a partir de diferentes perspectivas — como desempenho individual, evolução tática, impacto por posição, comparações entre eras da NBA, entre outras possibilidades. O potencial de aprofundamento é vasto e abre caminho para investigações ainda mais sofisticadas sobre o esporte.

## 6.2 Análise da qualidade da base de dados utilizada

Durante o processo de análise e preparação dos dados, foram identificados alguns problemas relacionados à qualidade e consistência das informações presentes nos datasets utilizados. Entre os principais pontos, destacam-se:

1. **Presença significativa de dados nulos:** Foi constatada uma grande quantidade de valores ausentes, principalmente nas temporadas mais antigas da NBA. Essa limitação reflete o histórico de registro incompleto ou impreciso das estatísticas durante os primeiros anos da liga, o que pode impactar diretamente em análises comparativas ao longo do tempo.

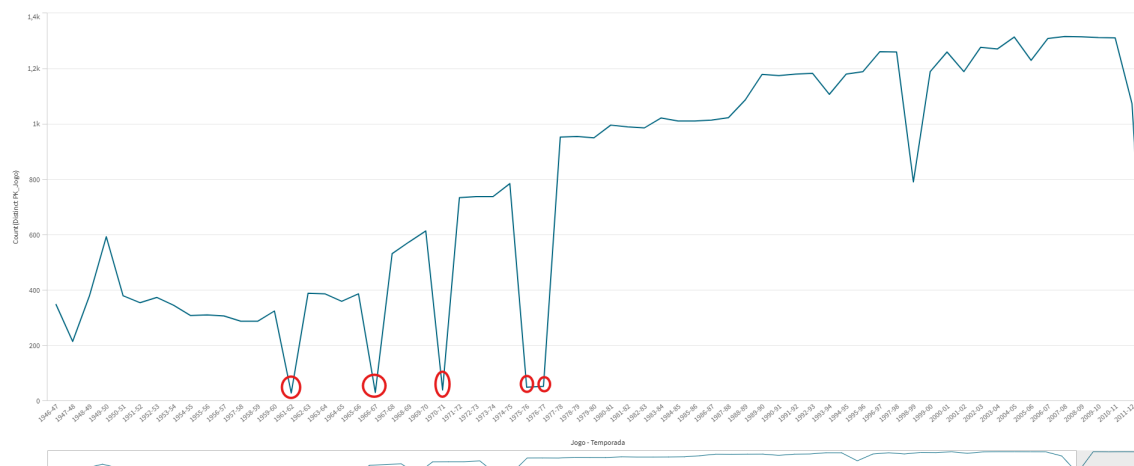
**Figura 21 – Valores Nulos - NBA Dados Históricos**

Times (Até 2016-17)													NBA < >
# Desempenho Times/Temporada													
Time / Temporada	Q1	PTS	PTS/G	AST	AST/G	STL	STL/G	REB	REB/G	Derrotas	Vitórias	Jogo - Temporada	Q1
Totais		2.332.558	107,17	4.446	22	664	8	15.663	53,64	9700	12065		
Boston Celtics/1949-50		5.419	79,89	0	-	0	-	0	-	46	22	1949-50	
New York Knicks/1949-50		5.024	81,15	0	-	0	-	0	-	30	42	1949-50	
Denver Nuggets/1949-50		2.044	78,62	0	-	0	-	0	-	16	10	1949-50	
Philadelphia Warriors/1949-50		2.813	76,85	0	-	0	-	0	-	16	18	1949-50	
Waterloo Hawks/1949-50		2.546	79,56	0	-	0	-	0	-	15	17	1949-50	
St. Louis Bombers/1949-50		2.467	74,76	0	-	0	-	0	-	14	19	1949-50	
Washington Capitols/1949-50		2.871	79,75	0	-	0	-	0	-	14	22	1949-50	
Tri-Cities Blackhawks/1949-50		3.196	85,57	0	-	0	-	0	-	13	24	1949-50	
Anderson Packers/1949-50		3.368	86,63	0	-	0	-	0	-	10	28	1949-50	
Indianapolis Olympians/1949-50		3.359	88,39	0	-	0	-	0	-	8	30	1949-50	
Chicago Stags/1949-50		2.590	78,48	0	-	0	-	0	-	7	26	1949-50	
Rochester Royals/1949-50		3.238	87,51	0	-	0	-	0	-	3	34	1949-50	
Syracuse Nationals/1949-50		3.738	86,93	0	-	0	-	0	-	2	41	1949-50	
Baltimore Bullets/1949-50		500	83,33	0	-	0	-	0	-	1	5	1949-50	
Minneapolis Lakers/1949-50		3.555	88,68	0	-	0	-	0	-	1	39	1949-50	
New York Knicks/1950-51		6.054	85,43	327	22	0	-	388	44,00	36	44	1950-51	
Boston Celtics/1950-51		6.026	84,90	26	20	0	-	0	-	32	39	1950-51	
Tri-Cities Blackhawks/1950-51		3.188	86,56	0	-	0	-	0	-	13	23	1950-51	
Indianapolis Olympians/1950-51		2.827	83,15	0	-	0	-	0	-	12	22	1950-51	
Washington Capitols/1950-51		1.596	84,00	0	-	0	-	0	-	12	7	1950-51	
Syracuse Nationals/1950-51		3.250	90,28	53	27	0	-	0	-	10	26	1950-51	
Philadelphia Warriors/1950-51		3.121	89,17	0	-	0	-	0	-	6	29	1950-51	
Rochester Royals/1950-51		3.807	90,84	110	28	0	-	223	55,79	4	36	1950-51	
Minneapolis Lakers/1950-51		3.284	84,21	0	-	0	-	0	-	4	33	1950-51	
Baltimore Bullets/1950-51		241	80,33	0	-	0	-	0	-	2	1	1950-51	
New York Knicks/1951-52		6.826	85,33	52	26	0	-	61	61,00	35	45	1951-52	
Boston Celtics/1951-52		6.316	91,54	18	18	0	-	0	-	29	40	1951-52	
Milwaukee Hawks/1951-52		2.029	75,15	0	-	0	-	0	-	13	14	1951-52	
Syracuse Nationals/1951-52		3.488	89,44	0	-	0	-	0	-	8	31	1951-52	

2. **Ausência de registros de times históricos:** Algumas equipes que fizeram parte da NBA em temporadas passadas não estavam presentes nos dados originais. Para garantir a integridade das análises, foi necessário realizar a inclusão manual dessas equipes com base em fontes confiáveis externas.
3. **Diferenças na nomenclatura de times:** Utilizando os dois datasets ocorreu que alguns times apresentavam variações na abreviação ou grafia do nome, dificultando os relacionamentos entre tabelas e exigindo ajustes padronizados para garantir a integridade das análises.

4. **Pouca quantidade de jogos registrados em determinadas temporadas:** Foi identificado que algumas temporadas específicas apresentam um número reduzido de registros de jogos, o que compromete a representatividade dos dados e dificulta a realização de análises precisas nesses períodos.

**Figura 22 – Ausência de Jogos em Determinadas Temporadas - NBA Dados Históricos**



Essas limitações exigiram um trabalho adicional de curadoria e tratamento de dados, reforçando a importância de uma etapa robusta de validação durante a construção de pipelines analíticos. Apesar desses desafios, os dados foram considerados suficientemente confiáveis e consistentes para a realização das análises propostas neste projeto.

## 7 CONCLUSÃO GERAL DO PROJETO

Este projeto teve como objetivo principal a construção de um pipeline de dados completo utilizando as plataformas Snowflake e Qlik Cloud, com foco na análise histórica da NBA desde 1950. Através da integração de diferentes datasets, tratamento e padronização das informações, modelagem dimensional e desenvolvimento de dashboards interativos, foi possível extrair insights relevantes sobre jogadores, times e tendências da liga ao longo das décadas.

A adoção do Snowflake como ferramenta de ingestão, armazenamento e processamento de dados mostrou-se eficiente, principalmente pela sua escalabilidade, facilidade de integração e estrutura clara em zonas (Landing, Curated e Consumption), que contribuíram para uma organização lógica e segura dos dados. Já o uso do Qlik Cloud permitiu a criação de visualizações ricas e dinâmicas, facilitando a exploração de informações complexas por meio de filtros, gráficos interativos e comparações detalhadas.

As perguntas levantadas no escopo do projeto puderam ser respondidas com base nas análises realizadas, permitindo observar diversos padrões presentes na NBA.

Apesar dos desafios encontrados — como a presença de dados nulos, inconsistências de nomenclatura e ausência de registros em temporadas iniciais — o projeto entregou uma base confiável e um ambiente analítico funcional que pode ser expandido futuramente com a inclusão de novas fontes, temporadas mais recentes ou até mesmo análises preditivas.

Por fim, este trabalho demonstra como a combinação entre boas práticas de engenharia de dados e ferramentas modernas de BI pode transformar dados brutos em informações valiosas para a tomada de decisão e para a compreensão mais profunda de contextos históricos e esportivos.

## 8 AUTOAVALIAÇÃO

Ao concluir este trabalho, percebo um avanço significativo nas minhas habilidades em engenharia de dados, especialmente no uso da plataforma Snowflake. A construção do pipeline completo, desde a ingestão até a transformação e disponibilização dos dados, me proporcionou uma visão prática e aprofundada sobre o funcionamento de ambientes analíticos modernos.

Além disso, estou bastante satisfeito com os resultados alcançados no Qlik. A criação de um dashboard interativo e bem estruturado demonstrou a eficácia das ferramentas utilizadas e mostrou que a solução poderia, inclusive, ser aplicada em contextos reais, como na própria NBA, para análises detalhadas de jogadores e times.

Outro ponto relevante foi o aprendizado sobre os dados estatísticos da NBA. O contato com essas informações permitiu não apenas a aplicação técnica dos conhecimentos adquiridos, mas também o aprofundamento no entendimento histórico e estratégico do esporte.

De forma geral, o projeto foi extremamente enriquecedor tanto no aspecto técnico quanto no pessoal, reforçando minha confiança na área de dados e despertando ainda mais interesse por projetos que unem tecnologia e esportes.