



**INSTITUTO POLITÉCNICO  
DO CÁVADO E DO AVE  
ESCOLA SUPERIOR  
DE TECNOLOGIA**

# Data Mart Implementation (P01)

DECISION SUPPORT SYSTEMS, 2023-24

Pedro Martins (23527), Luís Anjo (23528), Diogo Gomes (23893)

# 1. Índice

2. Introduction	4
3. Data sources	5
Resumo do Conteúdo da Base de Dados do Transfermarkt	6
4. Data profiling	7
Número de equipas por intervalos de competições	7
Capacidade média dos estádios	8
Número de jogos por época	9
Média de espectadores por jogo	10
Média de valores	11
Verificação de valores vazios	12
Competições sem país	13
5. Dimensional modelling	14
6. Matriz Data Warehouse	16
7. Querys SQL para responder às perguntas	17
Quem foram os melhores jogadores no último ano	17
Quais são os países onde se marcam mais golos	18
Quais os jogadores que são mais eficazes a marcar golos (por jogo)?	19
Quais as posições que têm mais percentagem de golos/assistências e cartões?	20
Quais os árbitros que dão mais vermelhos/amarelos (%)	20
Jogadores com mais golos/assistências/amarelos/vermelhos por pais	21
Minutos por cada avaliação do jogador	22
Treinadores com mais e menos golos marcados	23
8. Design of the dimensional data model	24
9. Data mart implementation	25
Dimensões	26
Resultados Obtidos:	35
10. Conclusão	38
11. Bibliography	39



## 2. Introdução

No âmbito da Unidade Curricular de Sistemas de Apoio à Decisão foi elaborado o presente relatório que foi proposto para o desenvolvimento e implementação de um Data Mart no qual será possível efetuar as pesquisas via *queries SQL* ao mesmo.

Para tal, procuramos analisar o conteúdo da base de dados utilizada, de forma a garantir que o Data Mart seja projetado de forma a proporcionar um acesso eficiente e preciso aos dados relevantes para as tomadas de decisão e preparar a modelação dimensional e o design do data model para a criação do Data Mart.

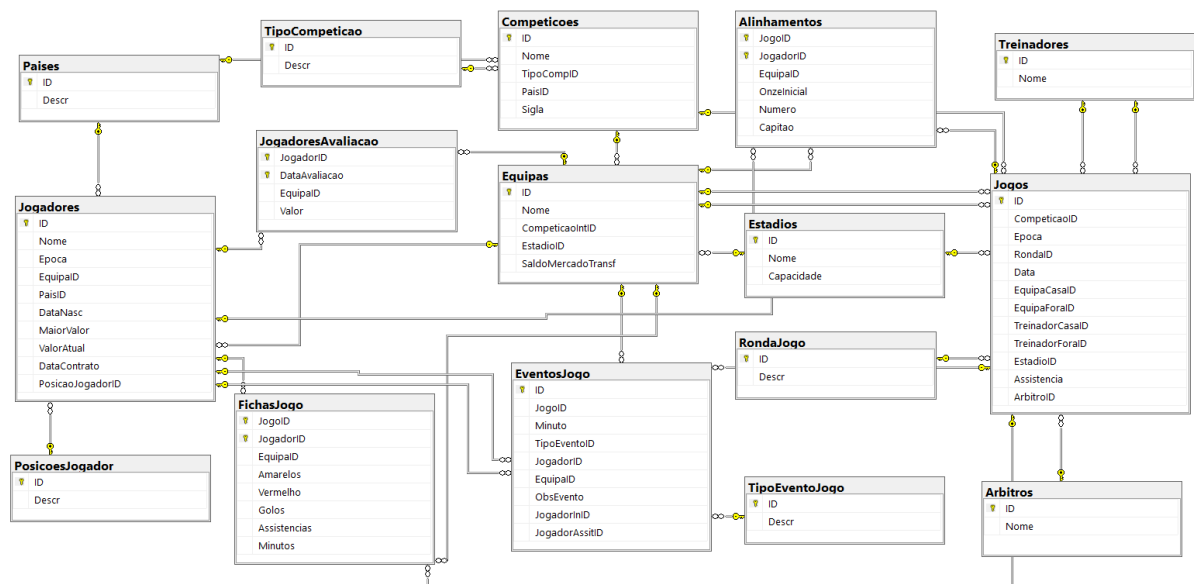
Ao longo deste relatório, serão abordados os passos seguidos durante o desenvolvimento do Data Mart, desde a conceção inicial até a sua implementação e disponibilização para utilização. Além disso, serão discutidas as metodologias e ferramentas utilizadas, bem como os desafios encontrados e as soluções adotadas ao longo do processo.

Esperamos assim que este relatório documente o trabalho realizado, mas também sirva como um recurso valioso para compreender os princípios e práticas envolvidos na criação de um Data Mart eficaz.

### 3. Data sources

A base de dados utilizada foi a base de dados “Transfermarkt” que contém dados relacionados com jogos, equipas e jogadores de futebol. Esta base de dados pertence ao Website alemão “Transfermarkt” que se dedica ao futebol e à avaliação de valores de mercado de vários jogadores.

Foi então carregado o diagrama ER para a base de dados utilizada, com a estrutura da mesma.



Foi também construída a tabela de Conteúdos, com a descrição, nome e número de entradas de dados de cada tabela da base de dados.

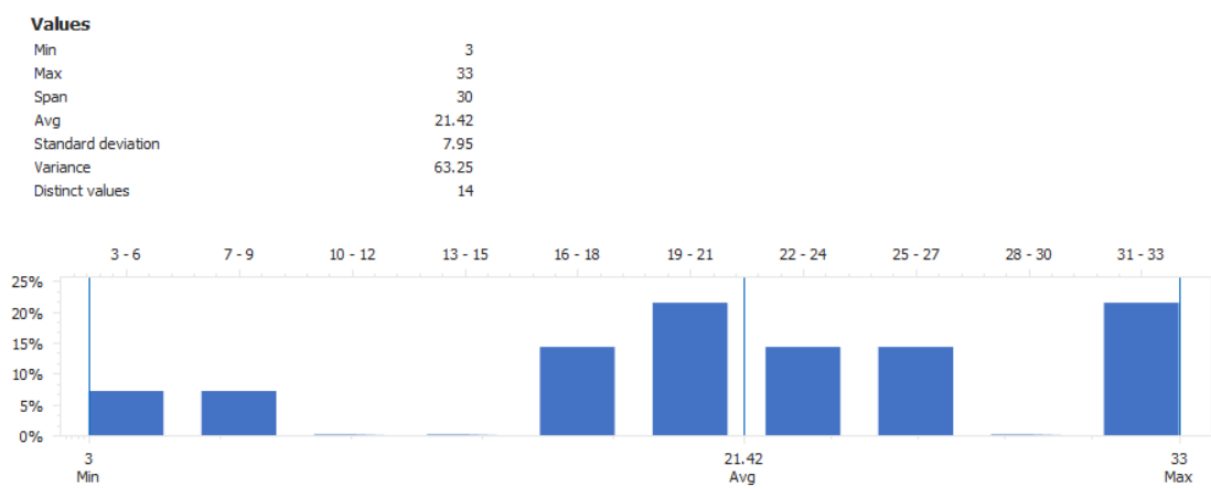
## Resumo do Conteúdo da Base de Dados do Transfermarkt

Objeto	Nome da Tabela	Nº Entradas
Jogadores	<i>Jogadores</i>	30 302
Posições dos Jogadores	<i>PosicoesJogador</i>	13
Avaliação dos Jogadores	<i>JogadoresAvaliacao</i>	440 571
Treinadores	<i>Treinadores</i>	5 553
Arbitros	<i>Arbitros</i>	2 335
Países	<i>Países</i>	180
Equipas	<i>Equipas</i>	426
Estádios	<i>Estádios</i>	2 280
Competições	<i>Competicoes</i>	43
Tipo das Competições	<i>TipoCompeticao</i>	11
Fichas dos Jogos	<i>FichasJogo</i>	1 432 321
Eventos dos Jogos	<i>EventosJogos</i>	518 713
Tipo de Eventos dos jogos	<i>TipoEventoJogo</i>	4
Alinhamentos das equipas nos jogos	<i>Alinhamentos</i>	69 909
Jogos	<i>Jogos</i>	51 499
Ronda da competição dos jogos	<i>RondaJogo</i>	116

## 4. Data profiling

Para efetuar o “data profiling” dos conteúdos da base de dados, foi utilizada a ferramenta “Dataedo”, que possibilitou extrair informação sobre a qualidade dos dados, através de métricas como média, valores máximos e mínimos, valores nulos, desvio, valores distintos, variância.

### Número de equipas por intervalos de competições



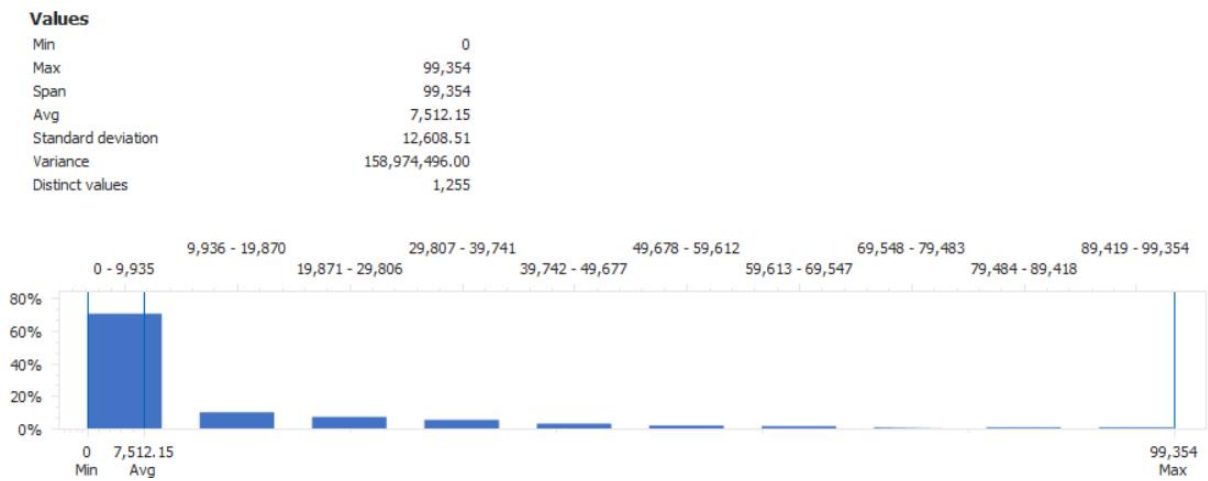
### Consulta SQL

	Intervalo	Contagem
1	3-6	29
2	7-9	28
3	16-18	58
4	19-21	104
5	22-24	66
6	25-27	53
7	31-33	88

### Código SQL

```
use Transfermarkt
SELECT
    Intervalo,
    COUNT(*) AS Contagem
FROM (
    SELECT
        CASE
            WHEN CompeticaoIntID BETWEEN 3 AND 6 THEN '3-6'
            WHEN CompeticaoIntID BETWEEN 7 AND 9 THEN '7-9'
            WHEN CompeticaoIntID BETWEEN 10 AND 12 THEN '10-12'
            WHEN CompeticaoIntID BETWEEN 13 AND 15 THEN '13-15'
            WHEN CompeticaoIntID BETWEEN 16 AND 18 THEN '16-18'
            WHEN CompeticaoIntID BETWEEN 19 AND 21 THEN '19-21'
            WHEN CompeticaoIntID BETWEEN 22 AND 24 THEN '22-24'
            WHEN CompeticaoIntID BETWEEN 25 AND 27 THEN '25-27'
            WHEN CompeticaoIntID BETWEEN 28 AND 30 THEN '28-30'
            WHEN CompeticaoIntID BETWEEN 31 AND 33 THEN '31-33'
        END AS Intervalo,
        CompeticaoIntID
    FROM dbo.Equipas
) AS T
GROUP BY Intervalo
ORDER BY MIN(CompeticaoIntID)
```

### Capacidade média dos estádios



### Consulta SQL

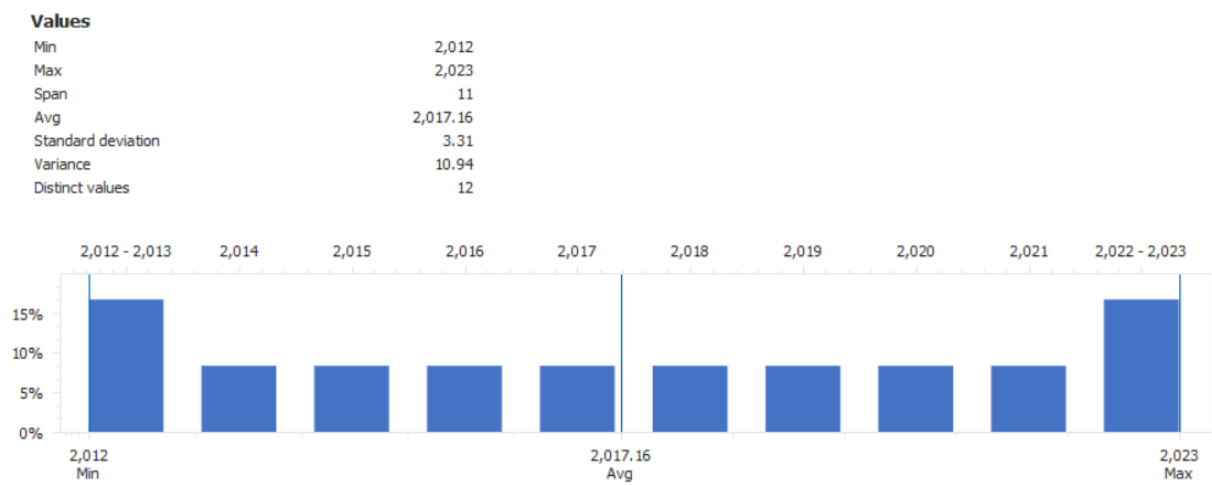
	Intervalo	Contagem
1	0 - 5K	1401
2	5K - 10K	493
3	10K - 20K	135
4	20K - 40K	161
5	40K - 60K	61
6	60K - 80K	24
7	+80K	5

### Código SQL

```
USE Transfermarkt
SELECT
    Intervalo,
    COUNT(*) AS Contagem
FROM (
    SELECT
        CASE
            WHEN Capacidade BETWEEN 0 AND 5000 THEN '0 - 5K'
            WHEN Capacidade BETWEEN 5000 AND 10000 THEN '5K - 10K'
            WHEN Capacidade BETWEEN 10000 AND 20000 THEN '10K - 20K'
            WHEN Capacidade BETWEEN 20000 AND 40000 THEN '20K - 40K'
            WHEN Capacidade BETWEEN 40000 AND 60000 THEN '40K - 60K'
            WHEN Capacidade BETWEEN 60000 AND 80000 THEN '60K - 80K'
            WHEN Capacidade > 80000 THEN '+80K'
        END AS Intervalo,
        Capacidade
    FROM dbo.Estadios
) AS T
WHERE Intervalo IS NOT NULL
GROUP BY Intervalo
ORDER BY MIN(Capacidade)
```



## Número de jogos por época



### Consulta SQL

	Intervalo	Contagem
1	2012-2013	9223
2	2014	4714
3	2015	4603
4	2016	4606
5	2017	4479
6	2018	4336
7	2019	4209
8	2020	4541
9	2021	4496
10	2022-2023	6292

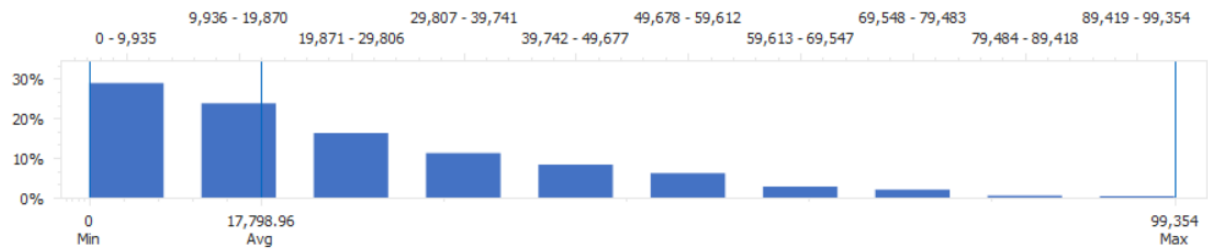
### Código SQL

```
SELECT
    Intervalo,
    COUNT(*) AS Contagem
FROM (
    SELECT
        CASE
            WHEN Epoca BETWEEN 2012 AND 2013 THEN '2012-2013'
            WHEN Epoca = 2014 THEN '2014'
            WHEN Epoca = 2015 THEN '2015'
            WHEN Epoca = 2016 THEN '2016'
            WHEN Epoca = 2017 THEN '2017'
            WHEN Epoca = 2018 THEN '2018'
            WHEN Epoca = 2019 THEN '2019'
            WHEN Epoca = 2020 THEN '2020'
            WHEN Epoca = 2021 THEN '2021'
            WHEN Epoca BETWEEN 2022 AND 2023 THEN '2022-2023'
        END AS Intervalo,
        Epoca
    FROM dbo.Jogos
) AS T
GROUP BY Intervalo
ORDER BY MIN(Epoca)
```

## Média de espectadores por jogo

### Values

Min	0
Max	99,354
Span	99,354
Avg	17,798.96
Standard deviation	18,348.25
Variance	336,658,112.00
Distinct values	25,196



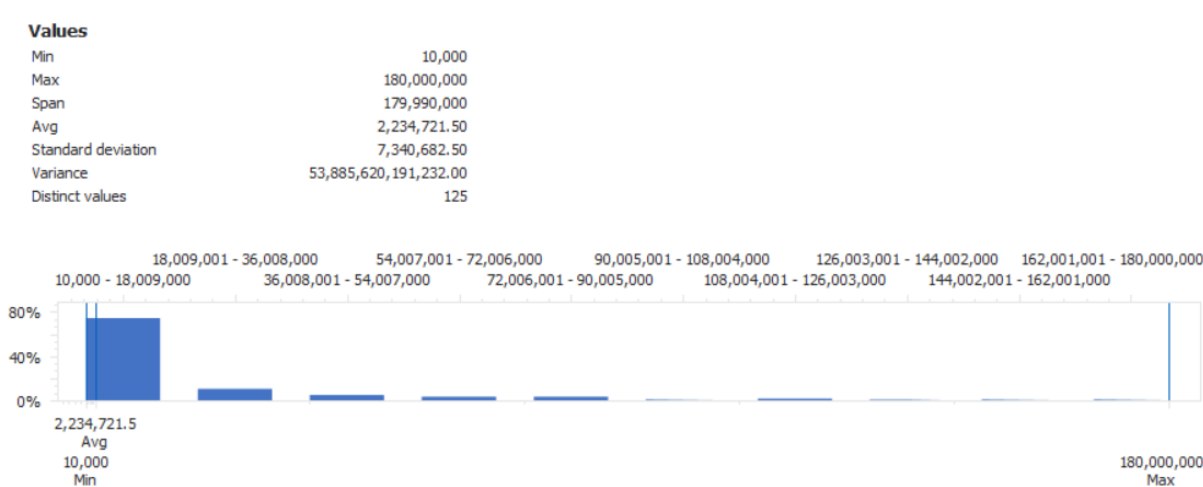
## Consulta SQL

Intervalo	Contagem
0 - 10K	23222
10K - 20K	10917
20K - 40K	10299
40K - 60K	5206
60K - 80K	1554
+80K	301

## Código SQL

```
USE Transfermarkt
SELECT
    Intervalo,
    COUNT(*) AS Contagem
FROM (
    SELECT
        CASE
            WHEN Assistencia BETWEEN 0 AND 10000 THEN '0 - 10K'
            WHEN Assistencia BETWEEN 10000 AND 20000 THEN '10K - 20K'
            WHEN Assistencia BETWEEN 20000 AND 40000 THEN '20K - 40K'
            WHEN Assistencia BETWEEN 40000 AND 60000 THEN '40K - 60K'
            WHEN Assistencia BETWEEN 60000 AND 80000 THEN '60K - 80K'
            WHEN Assistencia > 80000 THEN '+80K'
        END AS Intervalo,
        Assistencia
    FROM dbo.Jogos
) AS T
WHERE Intervalo IS NOT NULL
GROUP BY Intervalo
ORDER BY MIN(Assistencia)
```

## Média de valores



## Consulta SQL

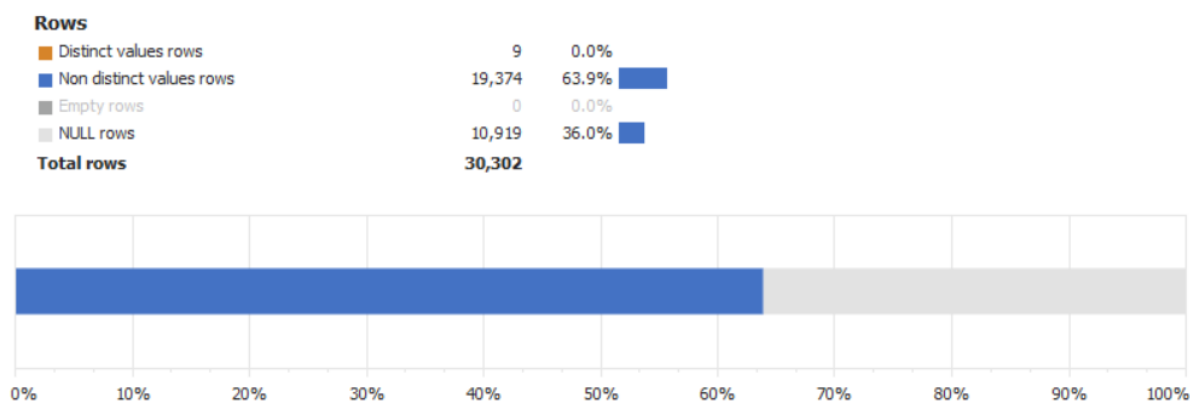
	Intervalo	Contagem
1	0 - 1M	14556
2	1M - 10M	3915
3	10M - 50M	825
4	50M - 70M	46
5	70M - 90M	27
6	90M - 100M	6
7	+100M	8

## Código SQL

```
USE Transfermarkt
SELECT
    Intervalo,
    COUNT(*) AS Contagem
FROM (
    SELECT
        CASE
            WHEN ValorAtual BETWEEN 0 AND 1000000 THEN '0 - 1M'
            WHEN ValorAtual BETWEEN 1000000 AND 10000000 THEN '1M - 10M'
            WHEN ValorAtual BETWEEN 10000000 AND 50000000 THEN '10M - 50M'
            WHEN ValorAtual BETWEEN 50000000 AND 70000000 THEN '50M - 70M'
            WHEN ValorAtual BETWEEN 70000000 AND 90000000 THEN '70M - 90M'
            WHEN ValorAtual BETWEEN 90000000 AND 100000000 THEN '90M - 100M'
            WHEN ValorAtual > 100000000 THEN '+100M'
        END AS Intervalo,
        ValorAtual
    FROM dbo.Jogadores
) AS T
WHERE Intervalo IS NOT NULL
GROUP BY Intervalo
ORDER BY MIN(ValorAtual)
```

# Verificação de valores vazios

Valor atual dos jogadores



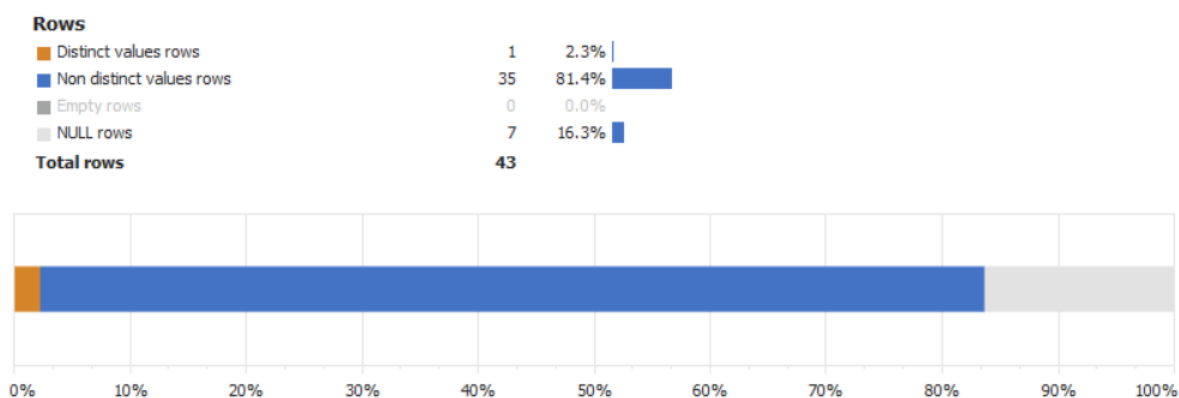
Consulta SQL

	row_type	row_count
1	Non Empty	19383
2	Null	10919

Código SQL

```
SELECT row_type,
COUNT(*) row_count
FROM
(
SELECT
CASE WHEN ValorAtual IS NULL THEN 'Null'
ELSE 'Non Empty'
END AS row_type
FROM dbo.Jogadores) rows
GROUP BY row_type
```

# Competições sem país



## Consulta SQL

	row_type	row_count
1	Non Empty	36
2	Null	7

## Código SQL

```
SELECT row_type,
COUNT(*) row_count
FROM
  (SELECT
    CASE WHEN PaisID IS NULL THEN 'Null'
    ELSE 'Non Empty'
    END AS row_type
    FROM dbo.Competicoes) rows
GROUP BY row_type
```

## 5. Dimensional modelling

Pretende-se desenvolver um Sistema de apoio à decisão para identificar padrões no futebol, estatísticas dos jogadores e das partidas. No sistema OLTP temos cerca de cinquenta mil dados de jogos em que podemos retirar parte destas informações. Para obtermos toda esta informação teremos de responder às seguintes questões:

- Quem foram os melhores jogadores no último ano?
- Quais são os países onde se marcam mais golos?
- Quais os jogadores que são mais eficazes a marcar golos (por jogo)?
- Quais as posições que têm mais percentagem de golos/assistências e cartões?
- Quais os árbitros que dão mais vermelhos/amarelos (%)?
- Jogadores com mais golos/assistências/amarelos/vermelhos por país?
- Minutos por cada avaliação do jogador?
- Treinadores com mais e menos golos marcados?

*Tabela 1: Dimensões*

Dimensão	Tabelas OLTP Source
dim_jogadoresPaíses	Jogadores, Países
dim_jogadoresPosicao	Jogadores, JogadoresPosicao
dim_jogadoresAvaliacao	Jogadores, JogadoresAvaliacao
dim_jogoArbTrei	Jogos, Arbitros, Treinadores
dim_date	Ficheiro CSV

Para responder às perguntas, foram analisadas as métricas as métricas e cálculos que deveriam ser utilizados.

Para exibir esses dados, elaboramos a seguinte tabela:

*Tabela 1: Métricas e Cálculos das Perguntas*

<b>Pergunta</b>	<b>Métricas e Cálculos Utilizadas</b>
Melhores jogadores no último ano	Golos e Assistencias do jogador; Algoritmo de pontuação (1 ponto para cada assistencia e 2 pontos para cada golo).
Países onde se marcam mais golos	Golos; Soma de golos por país.
Jogadores que são mais eficazes a marcar golos por jogo	Golos e Minutos; Contagem de jogos, soma de golos e media de golos por jogo.
Posições que têm mais percentagem de golos/assistências e cartões	Golos, Assistencias, Cartões Amarelos e Vermelhos; Soma de golos, assistencias e cartões e divisão para obter a percentagem.
Árbitros que dão mais cartões vermelhos/amarelos	Cartões Vermelhos e Amarelos; Contagem de jogos, soma de cartões e cálculo da media.
Jogadores com mais golos/assistências/amarelos/vermelhos por país	Golos, Assistencias, Cartões Amarelos e Vermelhos; Soma de golos, assistencias e cartões, contagem de países e jogadores e agregação por país.
Minutos por cada avaliação do jogador	Minutos e avaliações; Contagem de minutos por intervalo de datas de avaliação.
Treinadores com mais golos marcados	Golos; Contagem de soma de golos por treinador.

## 6. Matriz Data Warehouse

Com isto, foi também elaborada a Matriz do *Data Warehouse*, com as dimensões e os business processes.

Tabela 3: Matriz Data Warehouse

BUSINESS PROCESSES \ DIMENSIONS	jogadoresPaíses	jogadoresPaíses	jogadoresPosicao	jogoArbTrei	Dim_date
fichasJogo	X	X	X	X	X

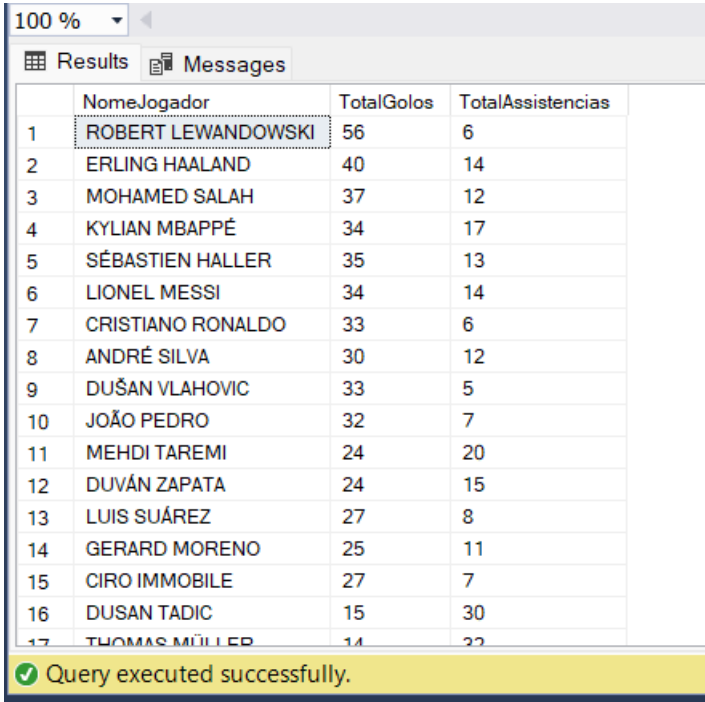
A matriz obtida é simples, uma vez que o esquema elaborado é um esquema estrela com apenas 5 dimensões e uma tabela factual.



## 7. Querys SQL para responder às perguntas

### Quem foram os melhores jogadores no último ano

```
SELECT
    NomeJogador,
    TotalGolos,
    TotalAssistencias
FROM (
    SELECT
        j.Nome AS NomeJogador,
        SUM(fj.Golos * 2 + fj.Assistencias) AS PontuacaoTotal,
        SUM(fj.Golos) AS TotalGolos,
        SUM(fj.Assistencias) AS TotalAssistencias
    FROM
        TransfermarktDW.dbo.dim_jogadoresPaises j
    INNER JOIN
        TransfermarktDW.dbo.fact_fichasJogo fj ON j.jogadorCode = fj.JogadorID
    INNER JOIN
        TransfermarktDW.dbo.dim_date d ON fj.dateKey = d.dateKey
    WHERE
        d.YearMonthNum BETWEEN 202101 AND 202112
    GROUP BY
        j.Nome
) AS PontuacaoJogador
ORDER BY
    PontuacaoTotal DESC;
```

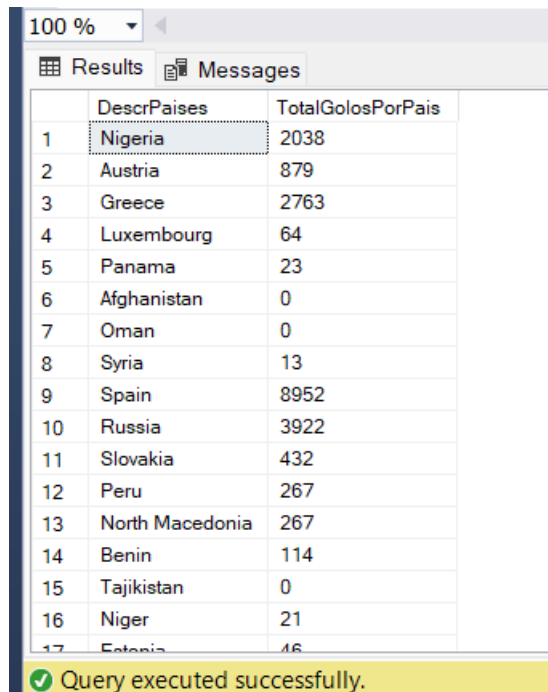


	NomeJogador	TotalGolos	TotalAssistencias
1	ROBERT LEWANDOWSKI	56	6
2	ERLING HAALAND	40	14
3	MOHAMED SALAH	37	12
4	KYLIAN MBAPPÉ	34	17
5	SÉBASTIEN HALLER	35	13
6	LIONEL MESSI	34	14
7	CRISTIANO RONALDO	33	6
8	ANDRÉ SILVA	30	12
9	DUŠAN VLAHOVIC	33	5
10	JOÃO PEDRO	32	7
11	MEHDI TAREMI	24	20
12	DUVÁN ZAPATA	24	15
13	LUIS SUÁREZ	27	8
14	GERARD MORENO	25	11
15	CIRO IMMOBILE	27	7
16	DUSAN TADIC	15	30
17	THOMAS MÜLLER	14	22

Query executed successfully.

## Quais são os países onde se marcam mais golos

```
SELECT
    dp.DescrPaises,
    SUM(ffj.Golos) AS TotalGolosPorPais
FROM
    [TransfermarktDW].[dbo].[fact_fichasJogo] ffj
JOIN
    [TransfermarktDW].[dbo].[dim_jogadoresPaises] dp ON ffj.jogadorPaisKey =
    dp.jogadorKey
GROUP BY
    dp.DescrPaises;
```



	DescrPaises	TotalGolosPorPais
1	Nigeria	2038
2	Austria	879
3	Greece	2763
4	Luxembourg	64
5	Panama	23
6	Afghanistan	0
7	Oman	0
8	Syria	13
9	Spain	8952
10	Russia	3922
11	Slovakia	432
12	Peru	267
13	North Macedonia	267
14	Benin	114
15	Tajikistan	0
16	Niger	21
17	Estonia	16

✓ Query executed successfully.

## Quais os jogadores que são mais eficazes a marcar golos (por jogo)?

```
WITH JogadoresGolos AS (
    SELECT
        fj.[JogadorID],
        COUNT(fj.[JogoID]) AS TotalJogos,
        SUM(fj.[Golos]) AS TotalGolos
    FROM
        [TransfermarktDW].[dbo].[fact_fichasJogo] fj
    GROUP BY
        fj.[JogadorID]
),
MédiaGolosPorJogo AS (
    SELECT
        jp.[jogadorKey],
        jp.[Nome],
        jg.TotalJogos,
        jg.TotalGolos,
        CASE
            WHEN jg.TotalJogos > 0 THEN jg.TotalGolos * 1.0 / jg.TotalJogos
            ELSE 0
        END AS MédiaGolosPorJogo
    FROM
        [TransfermarktDW].[dbo].[dim_jogadoresPaíses] jp
    LEFT JOIN
        JogadoresGolos jg ON jp.[jogadorCode] = jg.[JogadorID]
)
SELECT
    [Nome],
    TotalJogos,
    TotalGolos,
    MédiaGolosPorJogo
FROM
    MédiaGolosPorJogo
ORDER BY
    MédiaGolosPorJogo DESC;
```

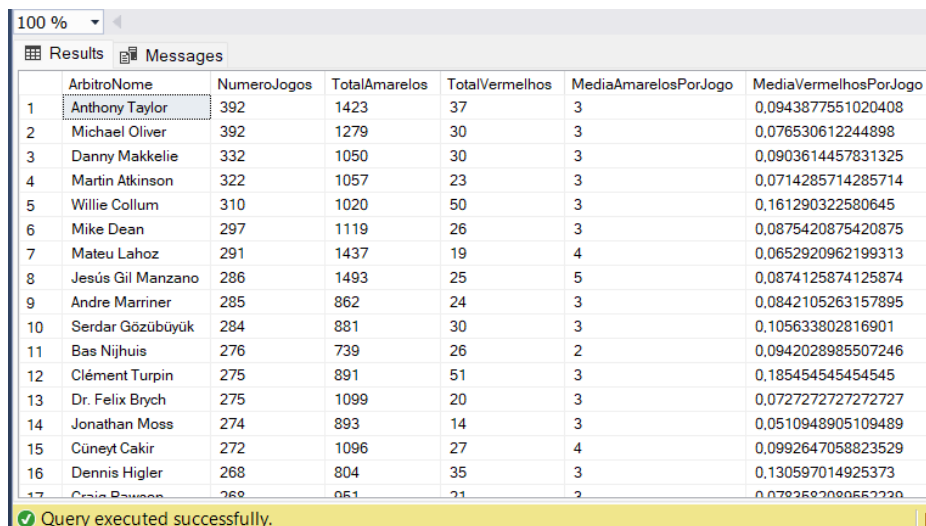
	Nome	TotalJogos	TotalGolos	MédiaGolosPorJogo
18	ERLING HAALAND	154	147	0.954545454545
19	CRISTIANO RONALDO	458	408	0.890829694323
20	LIONEL MESSI	504	434	0.861111111111
21	ROBERT LEWANDOWSKI	509	417	0.819253438113
22	MOHAMED AMOURA	11	9	0.818181818181
23	KYLIAN MBAPPÉ	293	215	0.733788395904
24	VIKTOR GYÖKERES	15	11	0.733333333333
25	RUBIN OKOTIE	15	11	0.733333333333
26	SAGIV JEHEZKEL	7	5	0.714285714285
27	ZLATAN IBRAHIMOVIC	274	192	0.700729927007
28	SERGIO AGÜERO	315	211	0.669841269841
29	HARRY KANE	398	266	0.668341708542
30	AARON BASTIAANS	3	2	0.666666666666
31	GUY MBENZA	3	2	0.666666666666
32	LUIS SUÁREZ	421	273	0.648456057007
33	SANTIAGO GIMENEZ	58	37	0.637931034482
34	AKOP ADAMS	11	7	0.636363636363

Query executed successfully.

Quais as posições que têm mais percentagem de golos/assistências e cartões?

Quais os árbitros que dão mais vermelhos/amarelos (%)

```
SELECT
    JAT.ArbitroNome,
    COUNT(DISTINCT FJ.JogoID) AS NumeroJogos,
    SUM(FJ.Amarelos) AS TotalAmarelos,
    SUM(CAST(FJ.Vermelho AS INT)) AS TotalVermelhos,
    SUM(FJ.Amarelos) / COUNT(DISTINCT FJ.JogoID) AS MediaAmarelosPorJogo,
    SUM(CAST(FJ.Vermelho AS FLOAT)) / COUNT(DISTINCT FJ.JogoID) AS MediaVermelhosPorJogo
FROM
    dbo.fact_fichasJogo FJ
JOIN
    dim_JogoArbTrei JAT ON FJ.jogoArbTreiKey = JAT.jogoKey
GROUP BY
    JAT.ArbitroNome
ORDER BY
    NumeroJogos Desc;
```



	ArbitroNome	NumeroJogos	TotalAmarelos	TotalVermelhos	MediaAmarelosPorJogo	MediaVermelhosPorJogo
1	Anthony Taylor	392	1423	37	3	0,0943877551020408
2	Michael Oliver	392	1279	30	3	0,076530612244898
3	Danny Makkellie	332	1050	30	3	0,0903614457831325
4	Martin Atkinson	322	1057	23	3	0,0714285714285714
5	Willie Collum	310	1020	50	3	0,161290322580645
6	Mike Dean	297	1119	26	3	0,0875420875420875
7	Mateu Lahoz	291	1437	19	4	0,0652920962199313
8	Jesús Gil Manzano	286	1493	25	5	0,0874125874125874
9	Andre Marriner	285	862	24	3	0,0842105263157895
10	Serdar Gözübüyük	284	881	30	3	0,105633802816901
11	Bas Nijhuis	276	739	26	2	0,0942028985507246
12	Clément Turpin	275	891	51	3	0,1854545454545454
13	Dr. Felix Brych	275	1099	20	3	0,0727272727272727
14	Jonathan Moss	274	893	14	3	0,0510948905109489
15	Cüneyt Çakır	272	1096	27	4	0,0992647058823529
16	Dennis Higler	268	804	35	3	0,130597014925373
17	Craig Dawson	268	951	21	2	0,0782592080552220

Query executed successfully.

## Jogadores com mais golos/assistências/amarelos/vermelhos por pais

```

SELECT
    est.DescrPaíses,
    MAX(CASE WHEN est.RankGolos = 1 THEN est.Nome ELSE NULL END) AS NomeGolos,
    MAX(CASE WHEN est.RankGolos = 1 THEN est.TotalGolos ELSE NULL END) AS TotalGolos,
    MAX(CASE WHEN est.RankAssistencias = 1 THEN est.Nome ELSE NULL END) AS
NomeAssistencias,
    MAX(CASE WHEN est.RankAssistencias = 1 THEN est.TotalAssistencias ELSE NULL END)
AS TotalAssistencias,
    MAX(CASE WHEN est.RankAmarelos = 1 THEN est.Nome ELSE NULL END) AS NomeAmarelos,
    MAX(CASE WHEN est.RankAmarelos = 1 THEN est.TotalAmarelos ELSE NULL END) AS
TotalAmarelos,
    MAX(CASE WHEN est.RankVermelhos = 1 THEN est.Nome ELSE NULL END) AS NomeVermelhos,
    MAX(CASE WHEN est.RankVermelhos = 1 THEN est.TotalVermelhos ELSE NULL END) AS
TotalVermelhos
FROM (
    SELECT
        d.DescrPaíses,
        d.Nome,
        SUM(f.Golos) AS TotalGolos,
        SUM(f.Assistencias) AS TotalAssistencias,
        SUM(f.Amarelos) AS TotalAmarelos,
        SUM(CONVERT(int, f.Vermelho)) AS TotalVermelhos,
        ROW_NUMBER() OVER(PARTITION BY d.DescrPaíses ORDER BY SUM(f.Golos) DESC) AS
RankGolos,
        ROW_NUMBER() OVER(PARTITION BY d.DescrPaíses ORDER BY SUM(f.Assistencias)
DESC) AS RankAssistencias,
        ROW_NUMBER() OVER(PARTITION BY d.DescrPaíses ORDER BY SUM(f.Amarelos) DESC) AS
RankAmarelos,
        ROW_NUMBER() OVER(PARTITION BY d.DescrPaíses ORDER BY SUM(CONVERT(int,
f.Vermelho)) DESC) AS RankVermelhos
    FROM dbo.fact_fichasJogo f
    JOIN dbo.dim_jogadoresPaíses d ON f.jogadorPaisKey = d.jogadorKey
    GROUP BY d.DescrPaíses, d.Nome
) AS est
GROUP BY est.DescrPaíses;
    
```

100 %

Results Messages

	DescrPaíses	NomeGolos	TotalGolos	NomeAssistencias	TotalAssistencias	NomeAmarelos	TotalAmarelos	NomeVermelhos	TotalVermelhos
1	Brazil	NEYMAR	203	NEYMAR	129	DANILO	177	MARCELO	7
2	French Guiana	SLOAN PRIVAT	27	LUDOVIC BAAL	12	YOANN SALMIER	13	LUDOVIC BAAL	2
3	Haiti	HERVÉ BAZILE	9	RÉGINAL GOREUX	13	RÉGINAL GOREUX	33	RÉGINAL GOREUX	1
4	Pakistan	MOHAMMAD ADNAN	2	MOHAMMAD ADNAN	4	NABIL ASLAM	3	NABIL ASLAM	0
5	Slovenia	JOSIP ILICIC	102	JOSIP ILICIC	60	JASMIN KURTIC	72	BOJAN JOKIC	2
6	Switzerland	BREEL EMOLO	49	XHERDAN SHAQIRI	39	GRANIT XHAKA	114	GRANIT XHAKA	5
7	Zambia	FASHION SAKALA	40	FASHION SAKALA	17	EVANS KANGWA	19	STOPPILA SUNZU	1
8	Zimbabwe	KNOWLEDGE MUSONA	40	QUINCY ANTIPAS	21	MARVELOUS NAKAMBA	43	TEENAGE HADEBE	2
9	Albania	SOKOL CIKALLESHI	41	SOKOL CIKALLESHI	15	FATJON ANDONI	75	MIGJEN BASHA	3
10	Bosnia-Herzegovina	EDIN DZEKO	181	EDIN VISCA	98	MIRALEM PJANIC	91	OGNJEN VRANJES	5
11	Congo	THIEVY BIFOUMA	32	THIEVY BIFOUMA	12	THIEVY BIFOUMA	35	DELVIN N'DINGA	1
12	Cyprus	PIEROS SOTIRIOU	29	KONSTANTINOS LAIFIS	9	ALEX GOGIC	33	KONSTANTINOS LAIFIS	2
13	Denmark	NICOLAI JØRGENSEN	107	CHRISTIAN ERIKSEN	120	THOMAS DELANEY	82	MADS AGESEN	4
14	Guinea-Bissau	MAMA BALDÉ	38	CARLOS MANÉ	29	PELÉ	41	ALFA SEMEDO	3
15	Guyana	TERELL ONDAAN	5	TERELL ONDAAN	5	TERELL ONDAAN	8	TERELL ONDAAN	0
16	Kyrgyzstan	VALERIY KICHIN	2	VALERIY KICHIN	0	VALERIY KICHIN	2	VALERIY KICHIN	0
17	Mexico	LINDSAY ROSE	8	LINDSAY ROSE	4	LINDSAY ROSE	22	LINDSAY ROSE	1

Query executed successfully.

LAPTOP-KN3F0NHQ (15.0 RTM) LAPTOP-KN3F0NHQ\diag

Query executed successfully.

LAPTOP-KN3F0NHQ (15.0 RTM) | LAPTOP-KN3F0NHQ\diag

## Minutos por cada avaliação do jogador

use TransfermarktDW

SELECT

```
jp.Nome AS NomeJogador,  
ja.Valor AS Avaliacao,  
ja.DataAvaliacao AS DataAvaliacao,  
SUM(fj.Minutos) AS TotalMinutosAteData
```

FROM

```
dbo.fact_fichasJogo fj
```

INNER JOIN

```
dbo.dim_jogadoresPaises jp ON fj.jogadorPaisKey = jp.jogadorKey
```

INNER JOIN

```
dbo.dim_jogadoresAvaliacao ja ON fj.jogadorID = ja.jogadorCode
```

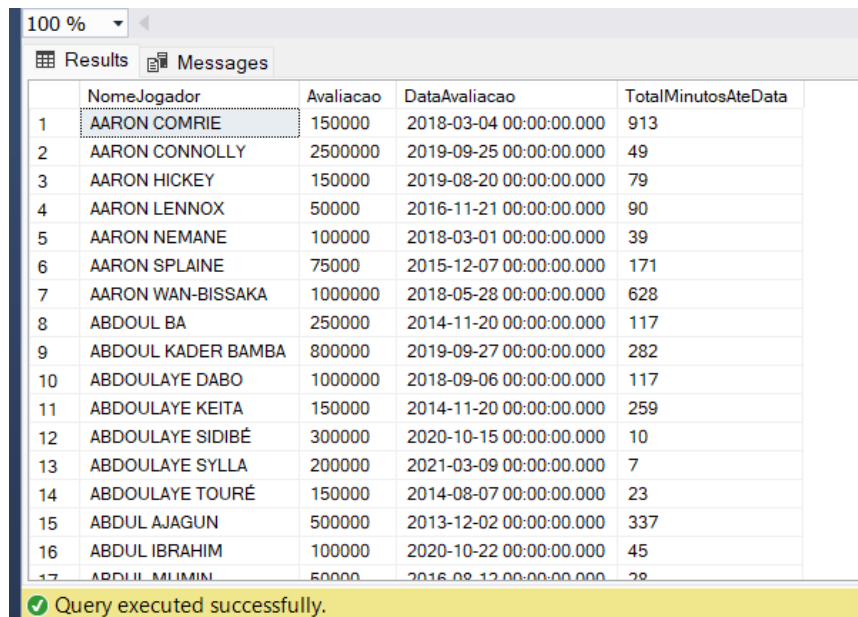
```
AND fj.dateKey <= CONVERT(int, REPLACE(CONVERT(varchar(10), ja.DataAvaliacao,  
112), '-', ''))
```

GROUP BY

```
jp.Nome,  
ja.Valor,  
ja.DataAvaliacao
```

ORDER BY

```
jp.Nome,  
ja.DataAvaliacao;
```



The screenshot shows a SQL Server Enterprise Manager interface. At the top, there's a zoom level of 100% and tabs for 'Results' and 'Messages'. The 'Results' tab is active, displaying a table with 17 rows and 5 columns. The columns are: 'NomeJogador', 'Avaliacao', 'DataAvaliacao', and 'TotalMinutosAteData'. The first row is highlighted. Below the table, a yellow status bar indicates 'Query executed successfully.'

	NomeJogador	Avaliacao	DataAvaliacao	TotalMinutosAteData
1	AARON COMRIE	150000	2018-03-04 00:00:00.000	913
2	AARON CONNOLLY	2500000	2019-09-25 00:00:00.000	49
3	AARON HICKEY	150000	2019-08-20 00:00:00.000	79
4	AARON LENNOX	50000	2016-11-21 00:00:00.000	90
5	AARON NEMANE	100000	2018-03-01 00:00:00.000	39
6	AARON SPLAINE	75000	2015-12-07 00:00:00.000	171
7	AARON WAN-BISSAKA	1000000	2018-05-28 00:00:00.000	628
8	ABDOUL BA	250000	2014-11-20 00:00:00.000	117
9	ABDOUL KADER BAMBA	800000	2019-09-27 00:00:00.000	282
10	ABDOULAYE DABO	1000000	2018-09-06 00:00:00.000	117
11	ABDOULAYE KEITA	150000	2014-11-20 00:00:00.000	259
12	ABDOULAYE SIDIBÉ	300000	2020-10-15 00:00:00.000	10
13	ABDOULAYE SYLLA	200000	2021-03-09 00:00:00.000	7
14	ABDOULAYE TOURÉ	150000	2014-08-07 00:00:00.000	23
15	ABDUL AJAGUN	500000	2013-12-02 00:00:00.000	337
16	ABDUL IBRAHIM	100000	2020-10-22 00:00:00.000	45
17	ABDUL MUMIN	50000	2016-08-12 00:00:00.000	28

## Treinadores com mais e menos golos marcados

```

SELECT
    NomeTreinador,
    COUNT(DISTINCT jogoKey) AS TotalJogos,
    SUM(Golos) AS Golos
FROM (
    SELECT
        dj.nomeTreinadorCasa AS nomeTreinador,
        ffj.Golos,
        dj.jogoKey
    FROM
        dim_JogoArbTrei dj
    JOIN
        fact_fichasJogo ffj ON dj.jogoKey = ffj.jogoArbTreiKey
    UNION ALL
    SELECT
        dj.nomeTreinadorFora AS nomeTreinador,
        ffj.Golos,
        dj.jogoKey
    FROM
        dim_JogoArbTrei dj
    JOIN
        fact_fichasJogo ffj ON dj.jogoKey = ffj.jogoArbTreiKey
) AS CombinedResults
GROUP BY
    NomeTreinador
ORDER BY
    TotalJogos DESC;

```

100 %

Results Messages

	NomeTreinador	TotalJogos	Golos
1	Diego Simeone	569	1293
2	Jürgen Klopp	552	1645
3	Pep Guardiola	537	1613
4	Unai Emery	514	1483
5	José Mourinho	474	1236
6	Carlo Ancelotti	467	1347
7	Gian Piero Gasperini	466	1310
8	Brendan Rodgers	457	1295
9	Massimiliano Allegri	455	1130
10	Stefano Pioli	439	1161
11	Christophe Galtier	428	984
12	Sérgio Conceição	427	1108
13	Mauricio Pochettino	415	1148
14	Manuel Pellegrini	409	1124
15	Ernesto Valverde	405	1102
16	Paulo Fonseca	401	1122
17	Mauricio Sari	305	1077

Query executed successfully.

## 8. Design of the dimensional data model

Para responder às perguntas colocadas anteriormente, definimos assim 5 dimensões, jogadoresAvaliacao, jogadoresPosicao, jogadoresPaíses, jogoArbTrei e dim\_date, e a tabela factual fichasJogo.

Com as tabelas de Dimensões e factual definidas, passamos assim ao desenvolvimento da modelação dimensional. Para isso foi elaborado o seguinte Diagrama de Entidade Relação com as tabelas de dimensões e factual e os respetivos campos.



A Tabela factual resultante do modelo, é considerada uma tabela transactional, já que cada entrada da tabela está relacionada com um jogo.

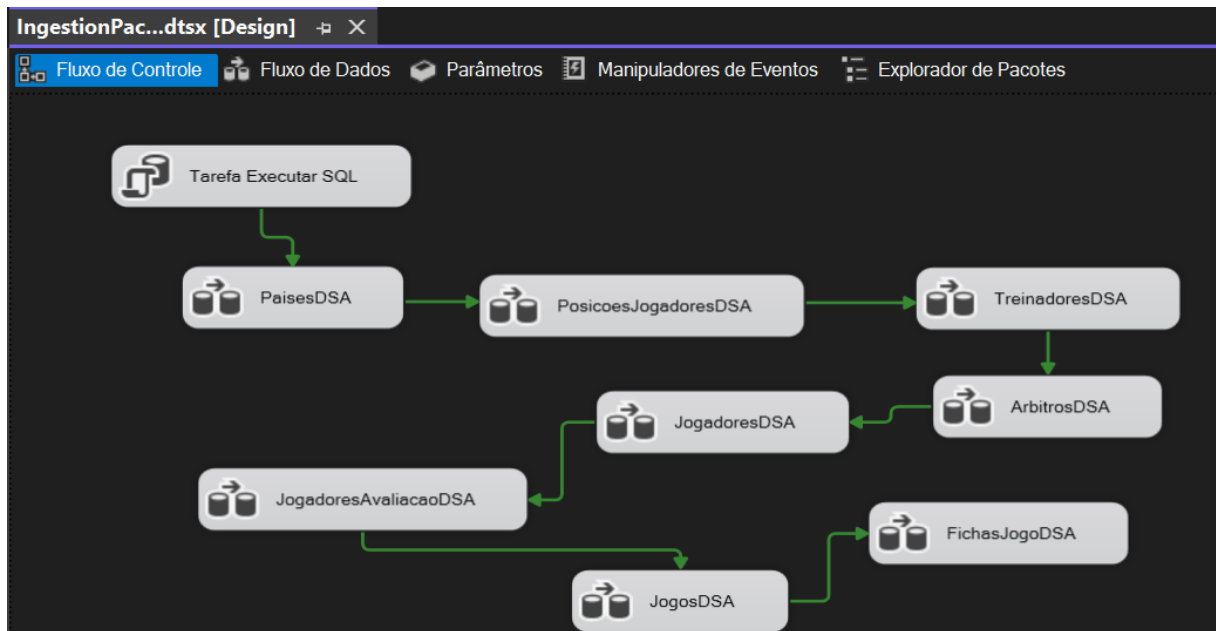
As métricas utilizadas na Tabela Factual representam as estatísticas dos jogos, neste caso serão os golos, assistencias, cartoes amarelos, cartoes vermelhos e minutos jogados, que conseguimos obter de forma direta, uma vez que foi utilizada a tabela fichasJogo do operacional, como base para a nossa tabela factual.

O nosso modelo não tem nenhum outrigger, já que temos um modelo estrela onde não temos nenhuma dimensão ligada a outra, e todas as nossas dimensões ficam ligadas diretamente à tabela factual.

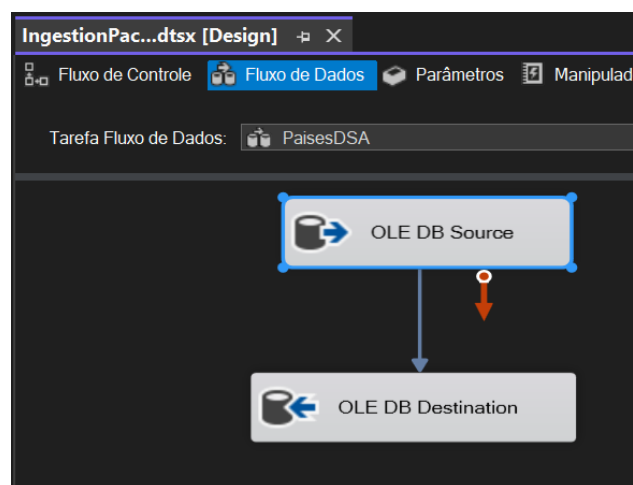


## 9. Data mart implementation

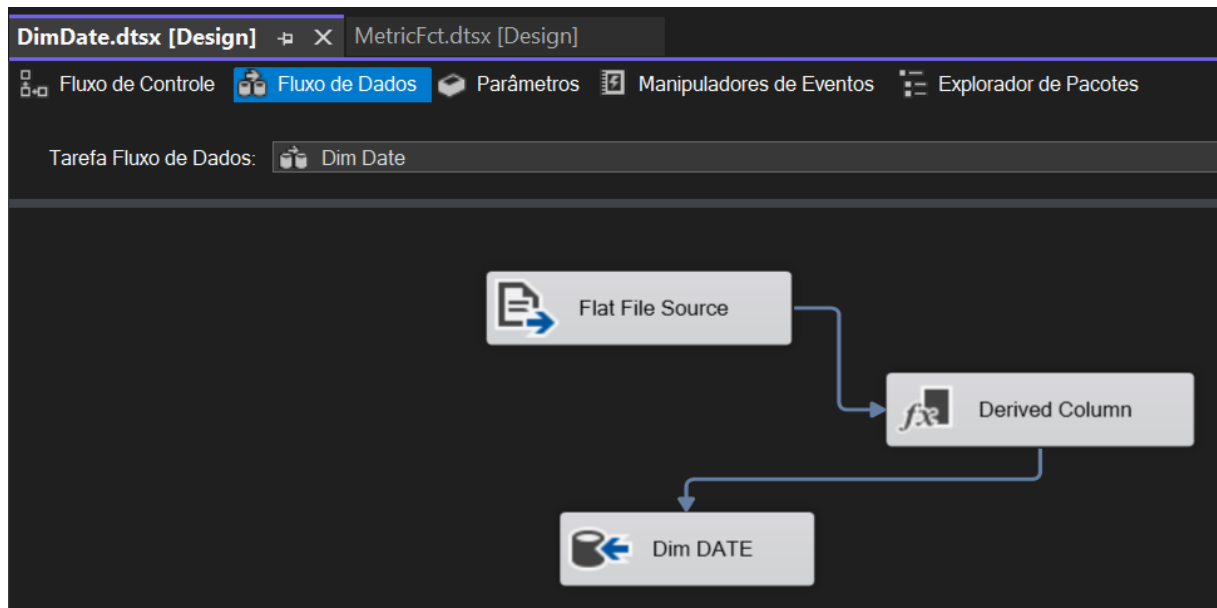
### Ingestão – Base de dados Operacional para DSA



Começamos Inicialmente por fazer a “ingestão” para as tabelas da base de dados DSA para conter os dados que serão posteriormente trabalhados, não usamos a base de dados de base pois essa não deve ser modificada. Fizemos no inicio o truncate das tabelas para retirar os dados e depois enviamos posteriormente para cada tabela os dados.



## Dimensões



Para obter todos os dados entre um determinado intervalo de datas começamos por arranjar um script para gerar uma tabela com os detalhes das datas entre um intervalo inicial e final:

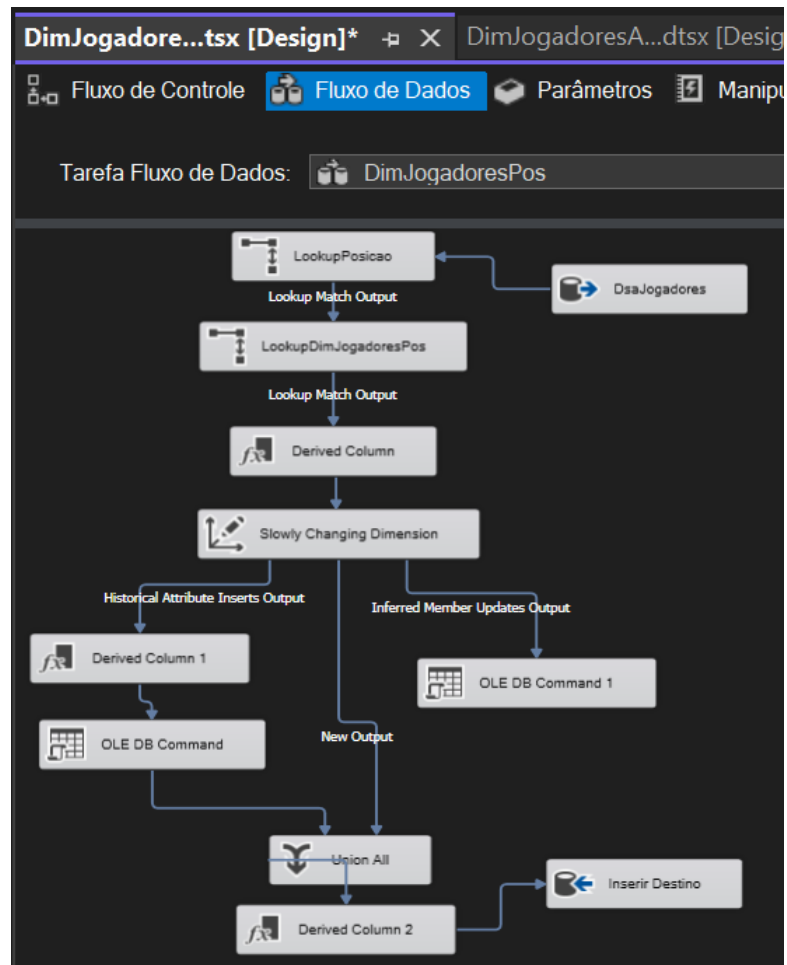
```
DECLARE @DataInicial DATE = '1967-01-01';
DECLARE @DataFinal DATE = '2024-12-01';

DECLARE @Datas TABLE ([Date] DATE);

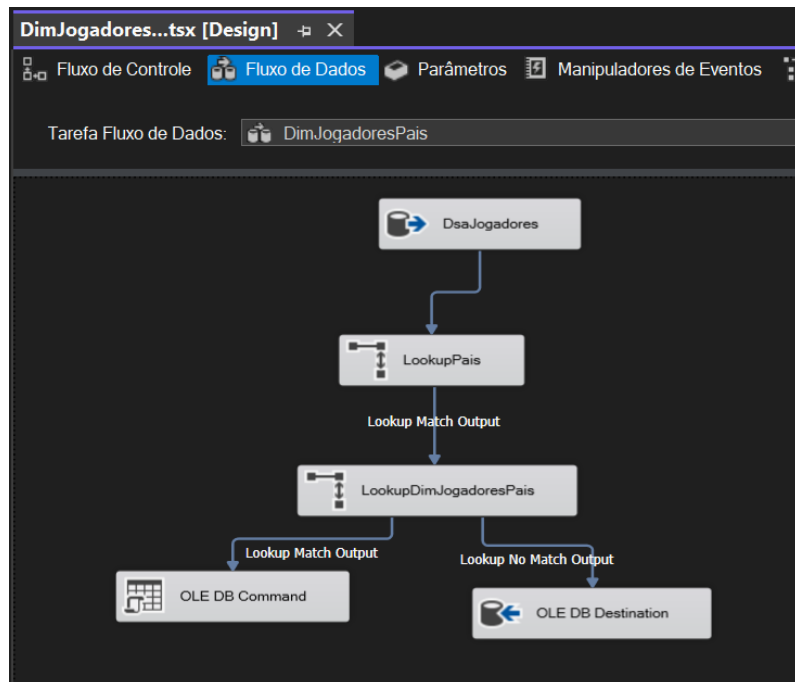
WHILE @DataInicial <= @DataFinal
BEGIN
    INSERT INTO @Datas VALUES (@DataInicial);
    SET @DataInicial = DATEADD(DAY, 1, @DataInicial);
END

INSERT INTO [dbo].[dim_date]
(
    [Date], [YearMonthNum], [Calendar_Quarter], [MonthNum],
    [MonthName], [MonthShortName], [WeekNum], [DayNumOfYear], [DayNumOfMonth],
    [DayNumOfWeek], [DayName], [DayShortName], [Quarter], [YearQuarterNum], [DayNumOfQuarter]
)
SELECT
    CONVERT(VARCHAR, [Date], 103) AS [Date],
    CONVERT(VARCHAR, YEAR([Date]) * 100 + MONTH([Date])) AS [YearMonthNum],
    'Q' + CONVERT(VARCHAR, DATEPART(QUARTER, [Date])) AS [Calendar_Quarter],
    CONVERT(VARCHAR, MONTH([Date])) AS [MonthNum],
    DATENAME(MONTH, [Date]) AS [MonthName],
    LEFT(DATENAME(MONTH, [Date]), 3) AS [MonthShortName],
    CONVERT(VARCHAR, DATEPART(WEEK, [Date])) AS [WeekNum],
    CONVERT(VARCHAR, DATEPART(DAYOFYEAR, [Date])) AS [DayNumOfYear],
    CONVERT(VARCHAR, DAY([Date])) AS [DayNumOfMonth],
    CONVERT(VARCHAR, DATEPART(WEEKDAY, [Date])) AS [DayNumOfWeek],
    DATENAME(WEEKDAY, [Date]) AS [DayName],
    LEFT(DATENAME(WEEKDAY, [Date]), 3) AS [DayShortName],
    'Q' + CONVERT(VARCHAR, DATEPART(QUARTER, [Date])) AS [Quarter],
    CONVERT(VARCHAR, YEAR([Date]) * 100 + DATEPART(QUARTER, [Date])) AS [YearQuarterNum],
    CONVERT(VARCHAR, DATEPART(DAYOFYEAR, [Date]) - DATEPART(DAYOFYEAR, DATEADD(QQ, 0, [Date])) + 1) AS [DayNumOfQuarter]
FROM
    @Datas;
```

Após a criação da tabela, extraímos os dados para um ficheiro csv, de forma a ter todas as datas em um ficheiro Excel para posteriormente carregar essas datas para a tabela “Dim\_date”.



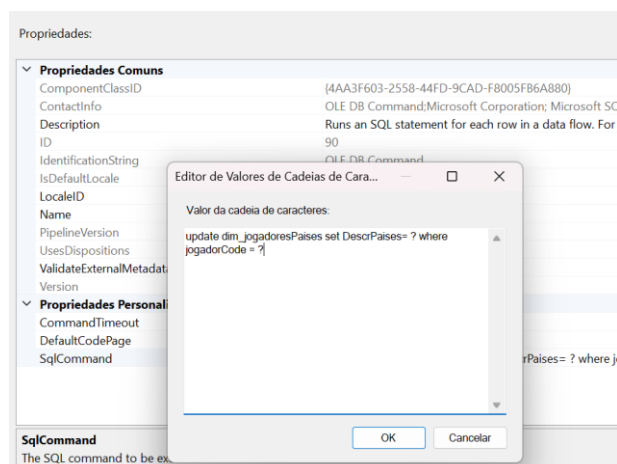
Para a criação da “Dim JogadoresPosicao” começamos por fazer o lookup com a posição do jogador para obtermos a mesma e depois usamos a slowly changing dimension para guardarmos o histórico de posições do jogador.

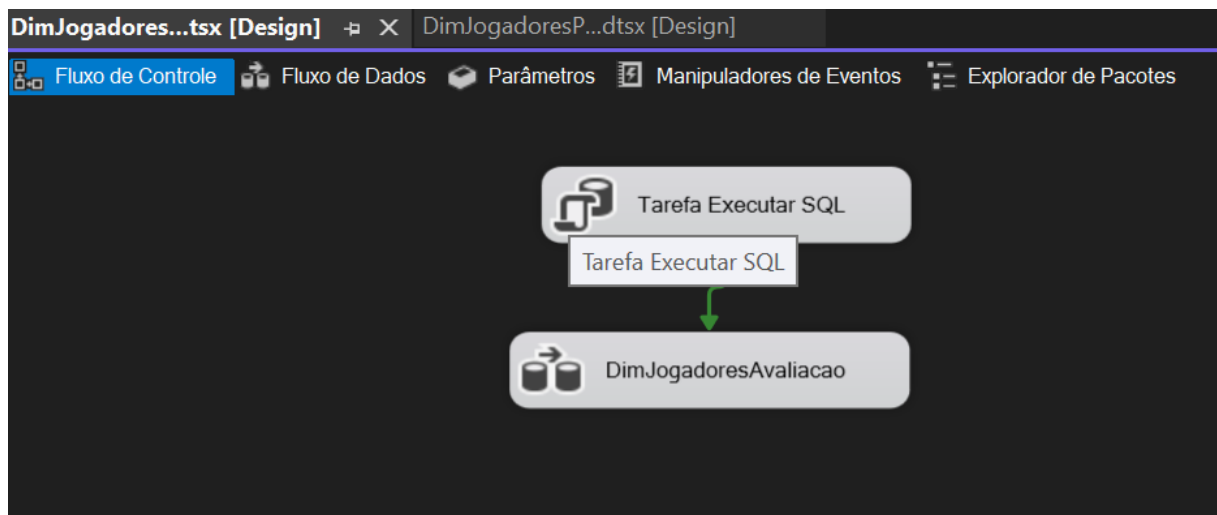


Para a criação da dimensão “DimJogadoresPaises” começamos por carregar os dados da tabela DSA dos jogadores. Após o carregamento dos dados fizemos um lookup com os dados da tabela DSA dos países de forma a juntar o jogador ao seu respetivo país com a ajuda do campo “PaisID” da tabela JogadoresDSA.

Após a inserção do nome do país para cada jogador fizemos um Lookup para verificar se o dado em questão já foi anteriormente adicionado. Caso o dado já tenha sido adicionado partirá para uma atualização do registo, caso este ainda não tenha sido adicionado partirá para a sua inserção na dimensão JogadoresPaises.

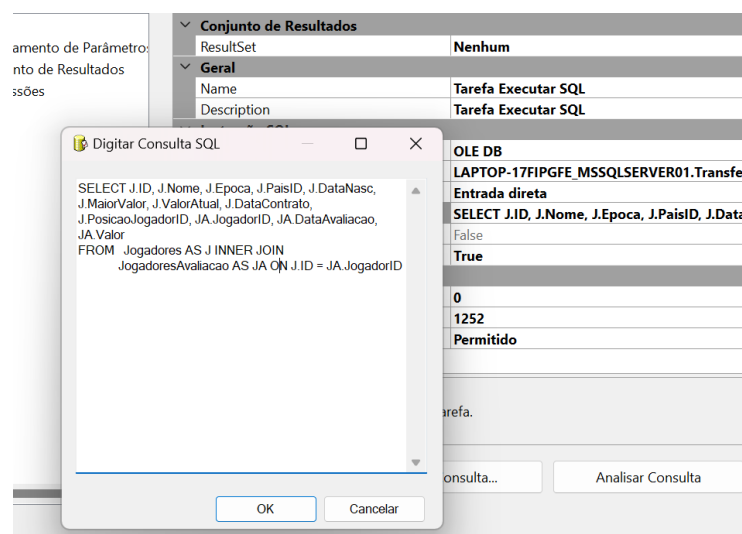
Aqui está o comando SQL para lidar com os dados que já foram inseridos anteriormente:

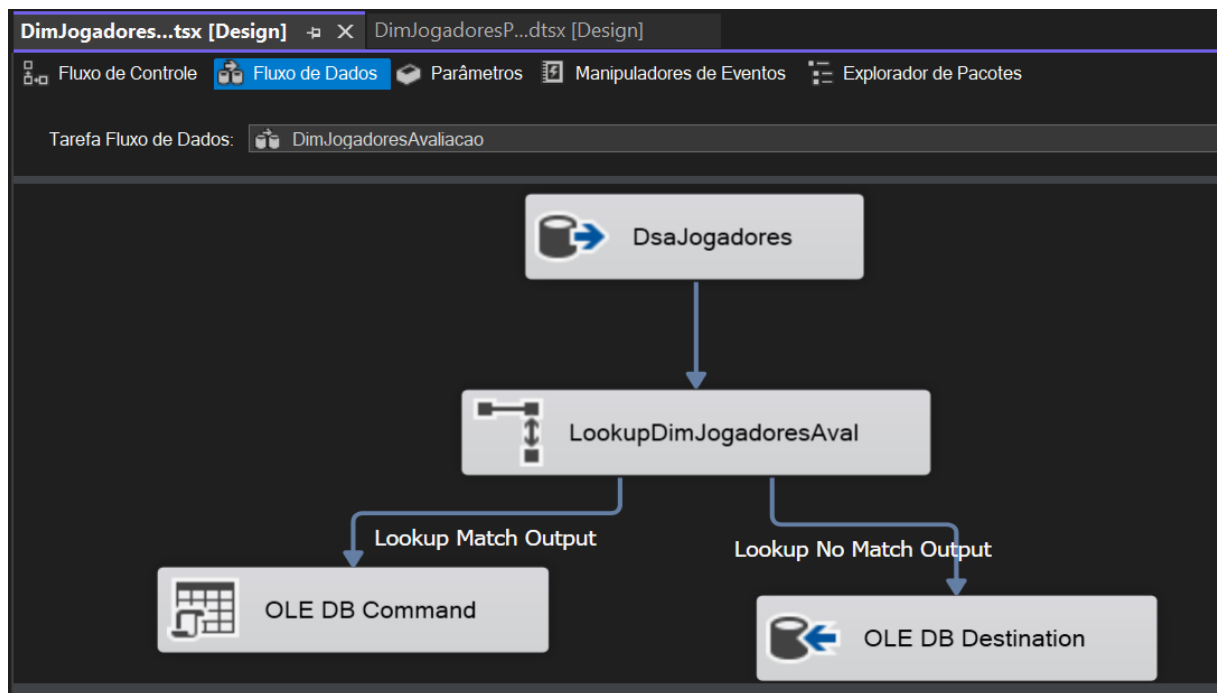




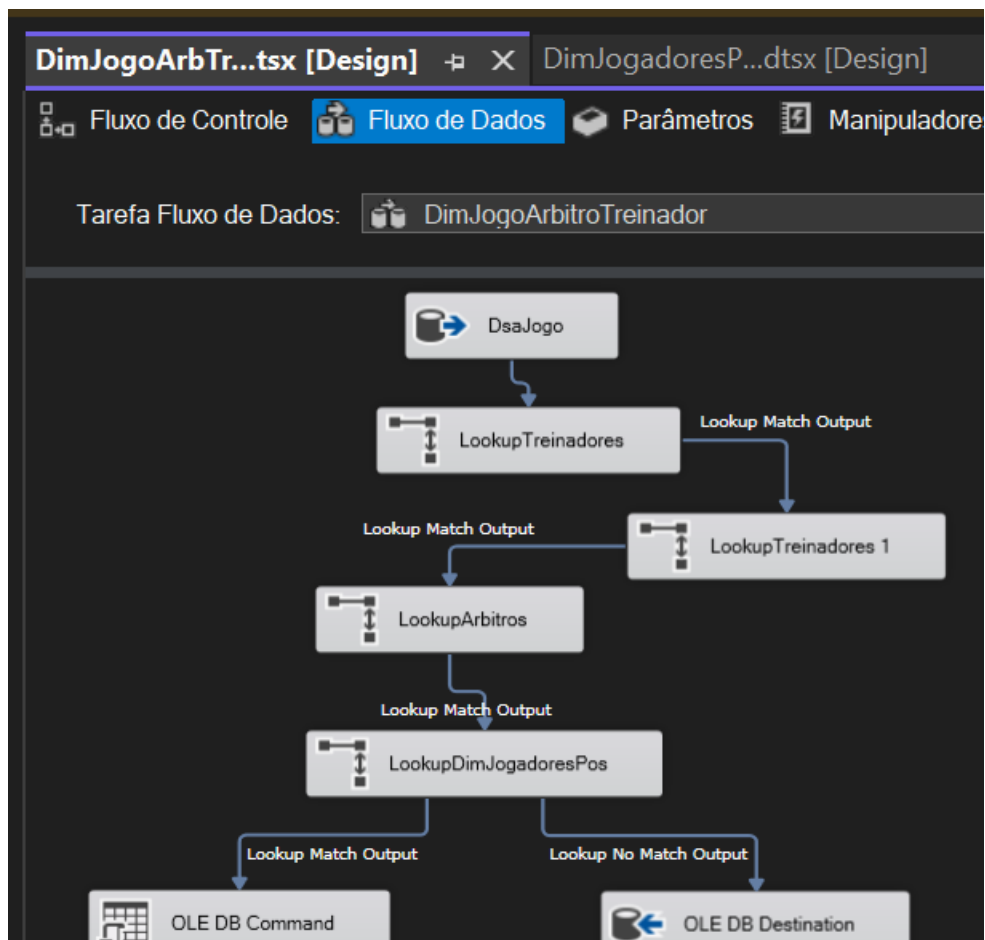
Para fazer a dimensão da avaliação dos jogadores tivemos que alterar um pouco a forma de fazer pois, como o jogador tinha um code e o mesmo aparecia várias vezes na tabela da avaliação, só trazia o primeiro que correspondia e a fazer o join conseguimos trazer todos antes de partir para o resto do processo.

Query Usada para fazer:

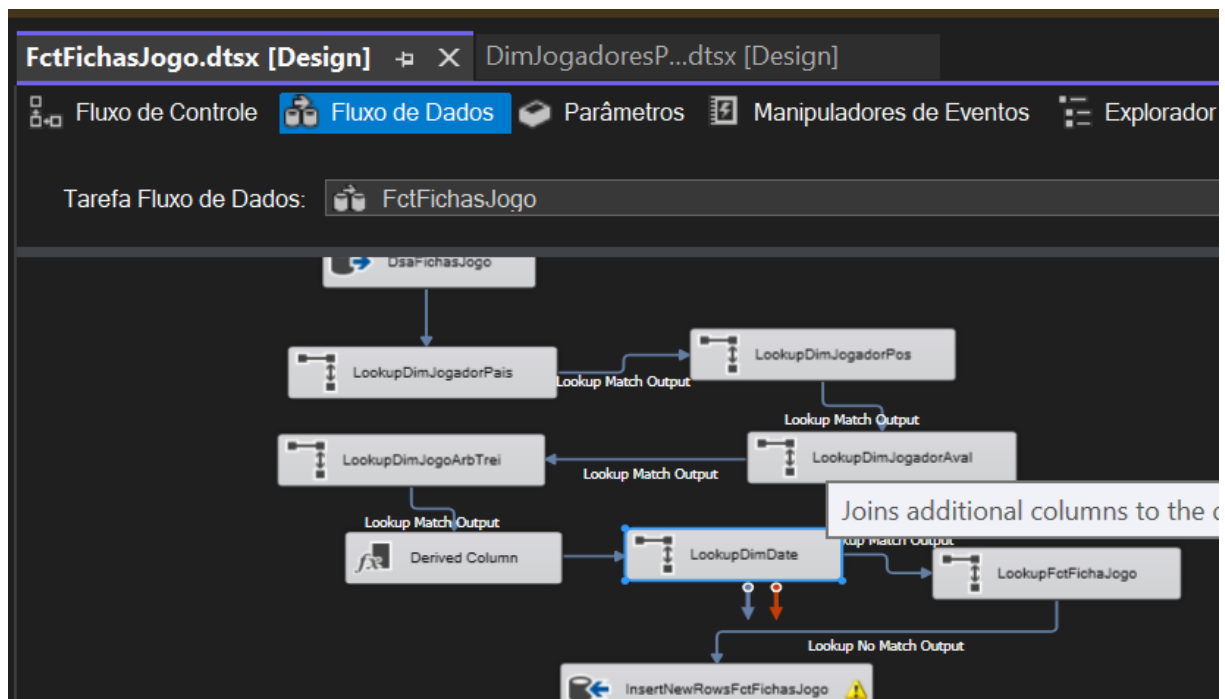




Para a dimensão “JogadoresAvaliacao” começamos por carregar os dados da tabela DSA dos jogadores. Após o carregamento dos dados fizemos um lookup com a dimensão.



A dimensão “Dim\_jogoArbTrei” consistirá na junção da tabela “Jogo” com a tabela “Treinadores” e “Arbitros”. Para a fazer começamos por carregar a tabela DSA com os dados relativamente ao jogo. Após o carregamento desta tabela fizemos primeiramente um lookup à tabela Treinadores para obter o nome do Treinador da casa. Após a inserção do campo com o Nome do treinador da equipa da casa, fizemos outro lookup à tabela Treinadores para obter desta vez o nome do treinador da equipa visitante. Já com os nomes dos treinadores dos 2 lados, fizemos um lookup à tabela dos Arbitros, desta vez para saber o nome do árbitro que arbitrou o jogo. Neste momento, temos a tabela Jogo com o nome de ambos os treinadores e o nome do árbitro. Para terminar a dimensão, fazemos um lookup para verificar se os dados já foram inseridos anteriormente. Caso estes já tenham sido inseridos, o package irá alterar o registo existente, caso estes sejam novos, irá inserir na dimensão “DimJogadoresPosicao”.



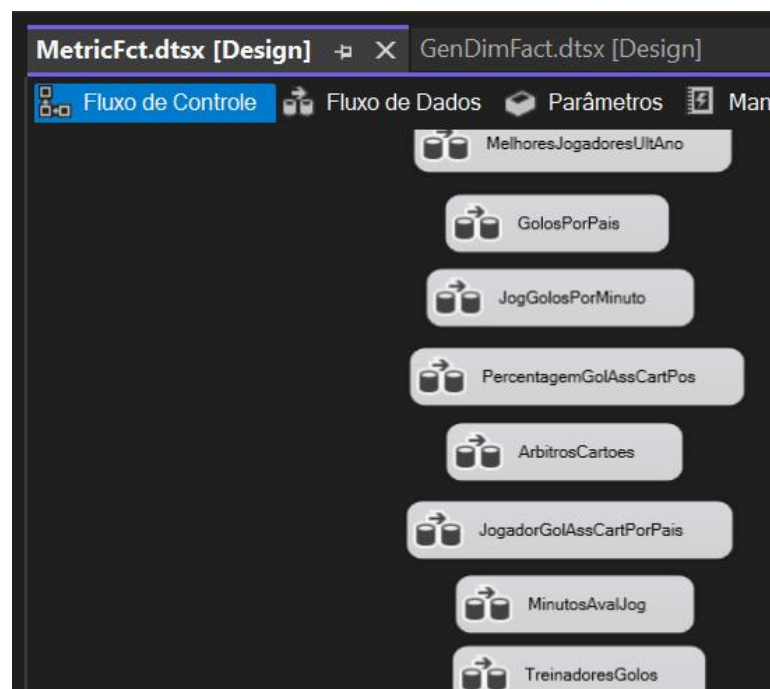
Após a criação de todas as dimensões pretendidas, partimos para a criação da nossa tabela factual. Começamos a nossa tabela factual com a tabela DSA “FichasJogo”, pois as nossas perguntas eram baseadas com as métricas que estavam contidas nessa tabela. Tendo os dados da tabela DSA “FichasJogo”, vamos fazer um lookup com cada dimensão criada anteriormente. Tendo realizado os lookups com as keys das dimensões, irá inserir os dados na tabela factual “fact\_fichasjogo”.

Para fazer o lookup com as datas tivemos que fazer primeiro a transformação na data que pretendíamos trazer com a dim date para ficar com o mesmo formato que a data presente na dim date. E assim obtivemos na factual a data com vários “formatos” para depois podermos responder às nossas perguntas.

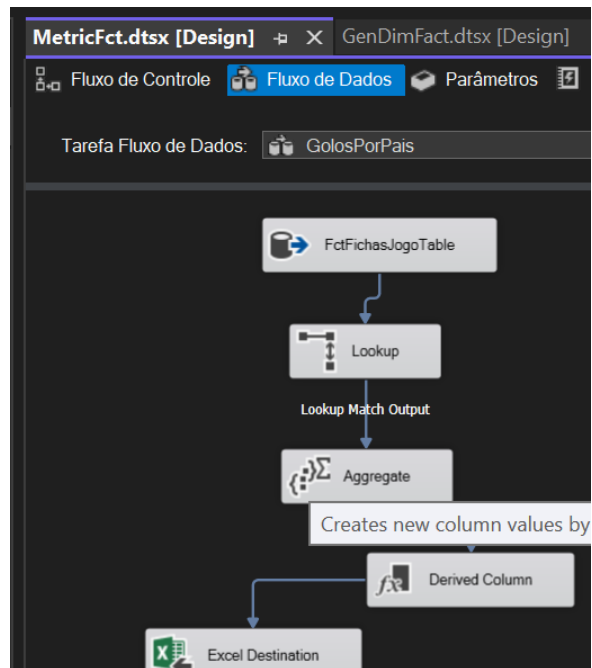




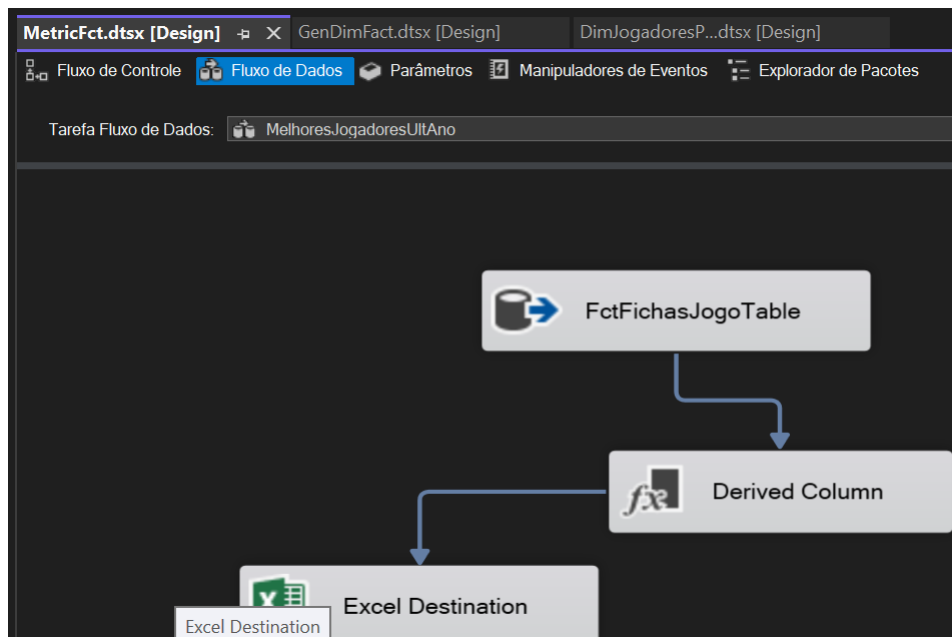
Depois de termos o ETL praticamente feito, já com dimensões e a tabela factural criada, criamos um package para compilar todos os packages das criações das dimensões juntamente com o package da criação da tabela factural. Este package será essencial para permitir a compilação de todas as dimensões incluindo a tabela factural sequencialmente.



Depois fizemos estes packages para fazer queries na tabela factural e obter as nossas métricas.



Esta query é um pouco diferente das outras visto que foram usados os componentes do integration services para fazer a mesma, as outras queries para responder às nossas perguntas foram feitas diretamente com queries sql.



Os outros packages foram feitos desta forma fizemos a query direta na tabela factual e depois tivemos que transformer todos os campos que obtivemos (métricas) para Unicode para conseguir colocar no excel.

### Resultados Obtidos:

			Nome Jogador	Total Golos	Total Assistencias	
			ROBERT LEWANDOWSKI	56	6	
			ERLING HAALAND	40	14	
			MOHAMED SALAH	37	12	
			KYLIAN MBAPPÉ	34	17	
			SÉBASTIEN HALLER	35	13	
			LIONEL MESSI	34	14	
			ANDRÉ SILVA	30	12	
			CRISTIANO RONALDO	33	6	
			JOÃO PEDRO	32	7	
			DUŠAN VLAHOVIC	33	5	
			MEHDI TAREMI	24	20	
			DUVÁN ZAPATA	24	15	
			LUIS SUÁREZ	27	8	
			GERARD MORENO	25	11	
			CIRO IMMOBILE	27	7	
			THOMAS MÜLLER	14	32	
			DUSAN TADIC	15	30	
			ANTOINE GRIEZMANN	22	14	
			WISSAM BEN YEDDER	25	6	
MelhoresJogadoresUltAno			GolosPorPais	GolosPorMinuto	PerGolAssCartPorPos	

	A	B	C
1	Pais	Golos	
2	Cameroon	1100	
3	Estonia	46	
4	Burundi	54	
5	Germany	5575	
6	Philippines	10	
7	Lebanon	4	
8	Togo	257	
9	Liechtenstein	2	
10	Ghana	1008	
11	Guyana	5	
12	Uganda	27	
13	Tunisia	442	
14	Angola	373	
< >		MelhoresJogadoresUltAno	GolosPorPais

o	GolosPorPais	<u>GolosPorMinuto</u>	PerGolAssCartPorPos	ArbitrosCartoes
---	--------------	-----------------------	---------------------	-----------------

10	GolosPorPais	GolosPorMinuto	<u>PerGolAssCartPorPos</u>	ArbitrosCartoes	JogadorGol
----	--------------	----------------	----------------------------	-----------------	------------

MelhoresJogadoresUltAno	GolosPorPais	GolosPorMinuto	PerGolAssCartPorPos	ArbitrosCartoes	Jogador
-------------------------	--------------	----------------	---------------------	-----------------	---------

País	Nome Jog	Golos	Total Gol	Nome Jog	Assistências	Total Assistências	Nome Jog	Amarelos	Total Amarelos	Nome Jog	Vermelhos	Total Vermelhos
Azerbaijã	ANATOLIY NURIYEV	10	PAVLO PASHAYEV	9	PAVLO PASHAYEV	41	KAMRAN AGAYEV	1				
Belgium	ROMELU LUKAKU	226	KEVIN DE BRUYNE	185	JAN VERTONGHEN	64	BIRGER VERSTRAETE	5				
China	LEI WU	9	LEI WU	3	LEI WU	6	LEI WU	0				
Chinese T.	TIM CHOW	3	TIM CHOW	1	TIM CHOW	16	TIM CHOW	2				
Czech Rep	PATRIK SCHICK	65	VLADIMIR DARIDA	40	VLADIMIR DARIDA	47	VLADIMIR DARIDA	2				
Dominican	MARIANO DÍAZ	26	MARIANO DÍAZ	5	MARIANO DÍAZ	15	HEINZ BARMETTLER	0				
Germany	THOMAS MÜLLER	173	THOMAS MÜLLER	193	DOMINIK KOHR	99	JÉRÔME BOATENG	4				
Guatemala	RUBIO RUBÍN	3	RUBIO RUBÍN	6	RUBIO RUBÍN	0	RUBIO RUBÍN	0				
Israel	LIOR REFAELOV	78	LIOR REFAELOV	65	LIOR REFAELOV	30	NIR BITTON	2				
Kazakhstan	BAKHTIYOR ZAYNUTDINOV	11	ALEKSANDR ZUEV	13	ALEKSANDR ZUEV	30	MARAT BYSTROV	2				
Russia	ARTEM DZYUBA	151	ARTEM DZYUBA	83	ANDREY SEMENOV	84	FEDOR KUDRYASHOV	6				
Rwanda	EDWIN OUON	2	EDWIN OUON	1	DJIHAD BIZIMANA	11	DJIHAD BIZIMANA	0				
Serbia	DUSAN TADIC	156	DUSAN TADIC	172	NEMANJA MATIC	88	STEFAN MITROVIC	5				
St. Kitts &	ROMAINE SAWYERS	0	ROMAINE SAWYERS	0	ROMAINE SAWYERS	3	ROMAINE SAWYERS	0				
Sweden	ZLATAN IBRAHIMOVIC	192	ZLATAN IBRAHIMOVIC	70	PONTUS WERNBLOOM	86	ZLATAN IBRAHIMOVIC	4				
Afghanistan	FARSHAD NOOR	0	FARSHAD NOOR	1	FARSHAD NOOR	2	FARSHAD NOOR	0				
Burundi	SAIDO BERAHINO	34	SAIDO BERAHINO	10	YOUSOUF NDAYISHI	19	YOUSOUF NDAYISHI	1				
Faroe Islan	HALLUR HANSSON	16	HALLUR HANSSON	10	HALLUR HANSSON	30	VILJORMUR DAVIDSEN	1				
France	KARIM BENZEMA	230	DIMITRI PAYET	120	BENJAMIN ANDRÉ	93	TÉJI SAVANIER	6				

GolosPorPaís GolosPorMinuto PerGolAssCartPorPos ArbitrosCartoes JogadorGolosAssCarPorPaís Treinadores

Nome do Treinador	Numero de Jogos	Numero de Golos
Diego Simeone	569	1293
Jürgen Klopp	552	1645
Pep Guardiola	537	1613
Unai Emery	514	1483
José Mourinho	474	1236
Carlo Ancelotti	467	1347
Gian Piero Gasperini	466	1310
Brendan Rodgers	457	1295
Massimiliano Allegri	455	1130
Stefano Pioli	439	1161
Christophe Galtier	428	984
Sérgio Conceição	427	1108
Mauricio Pochettino	415	1148
Manuel Pellegrini	409	1124
Ernesto Valverde	405	1102
Paulo Fonseca	401	1122
Maurizio Sarri	395	1077
Pedro Martins	390	962
Thomas Tuchel	388	1152

olAssCartPorPos ArbitrosCartoes JogadorGolosAssCarPorPaís TreinadoresGolos

## 10. Conclusão

Neste trabalho foi possível colocar em prática os conhecimentos obtidos na Unidade Curricular de Sistemas de Apoio à Decisão. Utilizamos conhecimentos adquiridos para implementar um Data Mart como foi previsto, através dos processos de análise, modelação e design dimensional, obtendo assim a estrutura dimensional que foi utilizada para desenvolver o Data Mart final.

Com o Data Mart elaborado, foi então possível responder a todas as perguntas efetuadas, através da manipulação e análise dos dados armazenados no Data Mart. Utilizando queries SQL conseguimos assim obter todos os dados pretendidos para responder às questões propostas e para uma futura análise.

O Modelo desenvolvido pode também ser escalado no futuro, uma vez que foi elaborado um esquema estrela com apenas uma tabela factual, e foram utilizadas apenas 7 tabelas da base de dados original, mas que pode eventualmente ser aumentado para cobrir outras perguntas de outros contextos diferentes.

No futuro é pretendido também utilizar estes dados obtidos para fazer uma análise mais extensa e detalhada aos mesmos, bem como a sua exibição organizada, que será facilitada através deste projeto, já que a sua implementação visa tornar todo o processo de obtenção de dados mais rápido e eficiente através da desnormalização da base de dados original.

Além disso, este trabalho demonstra a importância prática dos conceitos teóricos falados nesta unidade curricular, ao aplicar os conhecimentos num contexto real, sendo assim uma mais valia para valorizar o nosso entendimento sobre bases de dados e sobre os processos de obtenção dos dados da mesma.

## 11. Bibliography

### References

<https://dataedo.com/kb/query/sql-server/data-profiling-data-quality>

<https://www.thoughtspot.com/data-trends/data-modeling/slowly-changing-dimensions-in-data-warehouse>

Slides das aulas partilhados

## 12. Appendix A – Data description maps

Table 1: Data description map of Dim\_jogadoresPaíses

Name	Type of table	Nr. Records		Description				
Dim_JogadoresPaíses	Dimension	29759		Dimensão que contém os dados do jogador juntamente com o nome do seu país				
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example of values
jogadorKey	Surrogate key	int		Derivado			Surrogate Key	1, 2, 3, 4, 5
jogadorCode	Business key	int	1	Jogadores	ID	int		1, 2, 3, 4, 5
Nome	Nome do jogador	varchar	1	Jogadores	Nome	varchar		Pedro, Luis, Diogo
Epoca	Epoca do campeonato	int	1	Jogadores	Epoca	Int		2015, 2010, 2007
DataNasc	Data de Nascimento do jogador	datetime	1	Jogadores	DataNasc	datetime		1978-06-09 00:00:00.000, 1980-08-06 00:00:00.000, 1981-01-30 00:00:00.000
MaiorValor	Maior valor do jogador	int	1	Jogadores	MaiorValor	int		30000000, 24500000
ValorAtual	Valor atual do jogador	int	1	Jogadores	ValorAtual	int		250000, 1500000
DataContrato	Data do contrato do jogador	datetime	1	Jogadores	DataContrato	datetime		2024-06-30 00:00:00.000
DescrPaíses	Nome do país	varchar	1	Países	Descr	varchar		Belgium, Portugal



Table 2: Data description map of Dim\_jogadoresAvaliacao

Name	Type of table	Nr. Records		Description				
Dim_jogadoresAvaliacao	Dimension	440571		Dimensão que contém os dados do jogador juntamente com o valor e a data de avaliação				
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example of values
jogadorKey	Surrogate key	int		Derivado			Surrogate Key	1, 2, 3, 4, 5
jogadorCode	Business key	int	1	Jogadores	ID	int		1, 2, 3, 4, 5
Nome	Nome do jogador	varchar	1	Jogadores	Nome	varchar		Pedro, Luis, Diogo
Epoca	Epoca do campeonato	int	1	Jogadores	Epoca	Int		2015, 2010, 2007
DataNasc	Data de Nascimento do jogador	datetime	1	Jogadores	DataNasc	datetime		1978-06-09 00:00:00.000, 1980-08-06 00:00:00.000, 1981-01-30 00:00:00.000
MaiorValor	Maior valor do jogador	int	1	Jogadores	MaiorValor	int		30000000, 24500000
ValorAtual	Valor atual do jogador	int	1	Jogadores	ValorAtual	int		250000, 1500000
DataContrato	Data do contrato do jogador	datetime	1	Jogadores	DataContrato	datetime		2024-06-30 00:00:00.000
Valor	Valor do jogador	int	1	JogadoresAvaliacao	Valor	int		30000000, 24500000
DataAvaliacao	Data da avaliação	datetime	1	JogadoresAvaliacao	DataAvaliacao	datetime		2004-10-04 00:00:00.000, 2005-05-05 00:00:00.000, 2005-06-13 00:00:00.000

Table 2: Data description map of Dim\_jogadoresPosicao

Name	Type of table	Nr. Records	Description					
Dim_jogadoresPosicao	Dimension	30130	Dimensão que contém os dados do jogador juntamente com a posição do jogador					
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example of values
jogadorKey	Surrogate key	int		Derivado			Surrogate Key	1, 2, 3, 4, 5
jogadorCode	Business key	int	1	Jogadores	ID	int		1, 2, 3, 4, 5
Nome	Nome do jogador	varchar	1	Jogadores	Nome	varchar		Pedro, Luis, Diogo
Epoca	Epoca do campeonato	int	1	Jogadores	Epoca	Int		2015, 2010, 2007
DataNasc	Data de Nascimento do jogador	datetime	1	Jogadores	DataNasc	datetime		1978-06-09 00:00:00.000, 1980-08-06 00:00:00.000, 1981-01-30 00:00:00.000
MaiorValor	Maior valor do jogador	int	1	Jogadores	MaiorValor	int		30000000, 24500000
ValorAtual	Valor atual do jogador	int	1	Jogadores	ValorAtual	int		250000, 1500000
DataContrato	Data do contrato do jogador	datetime	1	Jogadores	DataContrato	datetime		2024-06-30 00:00:00.000
DescrPosicao	Posição do jogador	varchar	2	PosicoesJogador	Descr	Varchar		Goalkeeper, Centre-Forward, Centre-Back
StartDateKey	Data de criação	datetime	1	Dim_jogadoresPosicao	StartDateKey	datetime		1978-06-09 00:00:00.000, 1980-08-06 00:00:00.000, 1981-01-30 00:00:00.000
EndDateKey	Data de descontinuação	datetime	2	Dim_jogadoresPosicao	EndDateKey	datetime		1978-06-09 00:00:00.000, 1980-08-06 00:00:00.000, 1981-01-30 00:00:00.000
LastUpdateTime	Data da alteração	datetime	1	Dim_jogadoresPosicao	LastUpdateTime	datetime		1978-06-09 00:00:00.000, 1980-08-06 00:00:00.000

Table 4: Data description map of Dim\_jogoArbTrei

Name	Type of table	Nr. Records	Description					
Dim_ jogoArbTrei	Dimension	51222	Dimensão que contém os dados do jogo juntamente com o nome do árbitro					
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example of values
jogoKey	Surrogate key	int		Derivado			Surrogate Key	1, 2, 3, 4, 5
jogoCode	Business key	int	1	Jogos	ID	int		1, 2, 3, 4, 5
Epoca	Epoca do campeonato	int	1	Jogos	Epoca	Int		2015, 2010, 2007
Data	Data do jogo	datetime	1	Jogos	Data	datetime		1978-06-09 00:00:00.000, 1980-08-06 00:00:00.000, 1981-01-30 00:00:00.000
NomeTreinadorCasa	Nome do treinador da equipa da casa	varchar	1	Treinadores	Nome	varchar		José, Rui, Nelson
NomeTreinadorFora	Nome do treinador da equipa visitante	varchar	1	Treinadores	Nome	varchar		Roberto, Gil, Ricardo
Assistencia	Número de assistência no jogo	int	1	Jogos	Assistencia	int		619, 1437, 12808, 13872
ArbitroNome	Nome do árbitro	varchar	1	Arbitros	Nome	Varchar		Olegário Benquerença, Bruno Paixão, Paulo Baptista

Table 5: Data description map of Dim\_date

Name	Type of table	Nr. Records	Description					
Dim_date	Dimension	20820	Dimensão que contém os dados de uma data					
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example of values
dateKey	Surrogate key	int		Derivado			Surrogate Key	
Date	Data	varchar	1	Ficheiro CSV	Date	varchar		01/01/1967, 02/10/2000
YearMonthNum	Ano e mês	varchar	1	Ficheiro CSV	YearMonthNum	varchar		196701, 197507
Calendar_Quarter	Quarto do mês	varchar	1	Ficheiro CSV	Calendar_Quarter	varchar		Qtr 1, Qtr 2, Qtr 3
MonthNum	Número do mês	varchar	1	Ficheiro CSV	MonthNum	varchar		1,2,3,4
MonthName	Nome do mês	varchar	1	Ficheiro CSV	MonthName	varchar		January, February
MonthShortName	Nome curto do mês	varchar	1	Ficheiro CSV	MonthShortName	varchar		Jan, Feb, Mar
WeekNum	Número da semana	varchar	1	Ficheiro CSV	WeekNum	varchar		1, 2, 3, 4
DayNumOfYear	Número do dia do ano	varchar	1	Ficheiro CSV	DayNumOfYear	varchar		1, 2, 3, 4
DayNumOfMonth	Número do dia do mês	varchar	1	Ficheiro CSV	DayNumOfMonth	varchar		1, 2, 3, 4
DayNumOfWeek	Número do dia da semana	varchar	1	Ficheiro CSV	DayNumOfWeek	varchar		1, 2, 3, 4
DayName	Nome do dia	varchar	1	Ficheiro CSV	DayName	varchar		Sunday
DayShortName	Nome curto do dia	varchar	1	Ficheiro CSV	DayShortName	varchar		Sun
Quarter	Quarto	varchar	1	Ficheiro CSV	Quarter	varchar		Q1, Q2, Q3, Q4
YearQuarterNum	Quarto do ano	varchar	1	Ficheiro CSV	YearQuarterNum	varchar		196701
DayNumOfQuarter	Número do dia do quarto	varchar	1	Ficheiro CSV	DayNumOfQuarter	varchar		1, 2, 3, 4

Table 6: Data description map of fact\_fichasJogo

Name	Type of table	Nr. Records		Description				
fact_fichasJogo	Fact Table	1395917		Tabela factual que irá conter as dimensões e as métricas já existentes da tabela FichasJogo				
Target (Data mart)				Source (OLTP)				
Column	Description	Data type	SCD	Table	Column	Data type	ETL rules	Example of values
jogoID	Surrogate key	int		Derivado			Surrogate Key	1, 2, 3, 4, 5
jogadorID	Primary key	int	1	Derivado				1, 2, 3, 4, 5
jogadorPosKey	Foreign key	int	1	dim_jogadoresPosicao	jogadorKey	Int		1, 2, 3, 4, 5
jogadorPaisKey	Foreign key	int	1	Dim_jogadoresPaíses	jogadorKey	int		1, 2, 3, 4, 5
datekey	Foreign key	int	1	Dim_date	dateKey	int		1, 2, 3, 4, 5
jogoArbTreiKey	Foreign key	int	1	Dim_jogoArbTrei	jogoKey	int		1, 2, 3, 4, 5
Amarelos	Número de amarelos	int	1	FichasJogo	Amarelos	Int		0,1,2,3
Vermelho	Estado do vermelho	bit	1	FichasJogo	Vermelho	Bit		0,1
Golos	Número de golos	int	1	FichasJogo	Golos	Int		0,1,2,3
Assistencias	Número de assistências	int	1	FichasJogo	Assistencias	int		0,1,2,3
Minutos	Número de minutos jogados	int	1	FichasJogo	Minutos	int		12, 32, 90