



Soluções de Business Intelligence (BI) no Domínio do Futebol

TIAGO GONÇALVES BRAGANÇA

Junho de 2024

Soluções de Business Intelligence (BI) no Domínio do Futebol

Tiago Gonçalves Bragança

**Dissertação para obtenção do Grau de Mestre em
Engenharia Informática, Área de Especialização em
Sistemas de Informação e Conhecimento**

Orientador: Professor Doutor Paulo Oliveira

Co-orientador:

Júri:

Presidente:

Vogais:

Declaração de Integridade

Declaro ter conduzido este trabalho académico com integridade.

Não plagiei ou apliquei qualquer forma de uso indevido de informações ou falsificação de resultados ao longo do processo que levou à sua elaboração.

Portanto, o trabalho apresentado neste documento é original e de minha autoria, não tendo sido utilizado anteriormente para nenhum outro fim.

Declaro ainda que tenho pleno conhecimento do Código de Conduta Ética do P.PORTO.

ISEP, Porto, 29 de junho de 2024

Dedicatória

Dedico esta dissertação aos meus familiares, amigos e especialmente namorada pelo apoio incondicional. Sem eles nada disto seria possível.

Resumo

A análise de dados no futebol tem vindo a ser cada vez mais preponderante para o sucesso desportivo. Esta tem vindo a ser aplicada em diversos campos desde a monitorização de atletas, *scouting*¹, prevenção de lesões, previsões de comportamentos em campo, entre outros.

É do interesse de todos os profissionais de futebol terem a oportunidade de selecionar e consultar diversas informações que são relevantes para a sua profissão, especialmente para aqueles que precisam de tomar decisões constantemente. Neste sentido, a forma como a disponibilização de informação é realizada torna-se fulcral para a execução das suas tarefas, uma vez que estes necessitam que os dados sejam disponibilizados de forma clara e atempada.

No âmbito deste trabalho pretende-se desenvolver um sistema de informação de *Business Intelligence* (BI) de forma a colmatar as necessidades mencionadas, centralizando toda a informação sobre o desempenho dos jogadores e respetivas equipas nas diversas partidas de futebol. Desta forma, será possível auxiliar na tomada de melhores decisões desportivas para o futuro, sem a necessidade de realizar grandes alterações ao nível da estrutura do mesmo.

Neste está também incluído desenvolver diversos *dashboards* com indicadores sobre equipas, jogadores e ligas de forma a poder-se recolher conclusões acerca do aumento/diminuição de golos marcados e sofridos ao longo das temporadas, percentagem de golos marcados de bola parada, aumento/diminuição de influência de determinado jogador no decorrer de uma época desportiva, guarda-redes com menos golos sofridos, etc. Sendo um conjunto de indicadores de elevada importância na preparação de próximos encontros ou próximas épocas desportivas.

Por fim, para validar o produto desenvolvido, procede-se à realização de um inquérito dirigido a uma equipa técnica de futebol, onde foram avaliadas a usabilidade e a utilidade do sistema. De acordo com os resultados obtidos, o desenvolvimento da aplicação atingiu os objetivos definidos e demonstrou utilidade para os profissionais de futebol.

Palavras-chave: Futebol, Business Intelligence, Armazém de Dados, OLAP, Sports Analytics, PowerBI.

¹ *Scouting* – Processo definido pela análise e observação, cujo principal objetivo é obtenção do máximo de informação possível tanto a nível individual como coletivo. Termo muitas vezes utilizado na procura de potenciais jovens atletas promissores.

Abstract

Data analysis in football has become increasingly important for sporting success. It has been applied in various fields, from monitoring players, scouting, injury prevention and predicting behavior on the pitch, among others.

It is in the interest of all football professionals to have the opportunity of selecting and consulting several data that are relevant to their job, especially for those who need to make decisions constantly. In this sense, the way in which information is made available becomes central to the execution of their tasks, since they require timely and clear data availability.

The aim of this work is to develop a *Business Intelligence* (BI) information system to meet these needs, centralizing all the information on the sporting performance of players and their teams in several football matches. In this way, it will be possible to help make better sporting decisions for the future, without the need to make major changes to its structure.

It also includes the development of various dashboards with indicators on teams, players and leagues so that conclusions can be drawn about the increase/decrease in goals scored and conceded over the course of seasons, the percentage of goals scored from set pieces, the increase/decrease in a certain player's influence over the course of a season, goalkeepers with fewer goals conceded, etc. This is a set of highly important indicators when preparing for upcoming matches or sporting seasons.

Finally, to validate the product developed, a survey was carried out with a football coaching team, where the usability and usefulness of the system were assessed. According to the results obtained, the development of the application achieved the defined objectives and proved useful for football professionals.

Keywords: Football, Business Intelligence, Data Warehouse, OLAP, Sports Analytics, PowerBI.

Agradecimentos

A realização desta dissertação foi um dos obstáculos mais desafiantes na minha vida, na qual teve contribuições de diversas pessoas que me são importantes.

Primeiramente, quero agradecer aos meus avôs, país, irmãos e amigos que fazem parte da minha vida de forma ativa. Cada um teve a sua contribuição no meu percurso, de forma maior ou menor, mas igualmente importante.

Agradeço ao meu Orientador, Professor Doutor Paulo Oliveira pela confiança, paciência, constante disponibilidade no esclarecimento de dúvidas e orientação em todo o processo.

Um especial agradecimento à minha namorada, pela contínua motivação, dedicação e incentivo. Tendo sido a principal contribuinte na conclusão desta etapa da minha vida.

Quero também agradecer a toda a equipa técnica do clube Rans Nusantara FC que se disponibilizou para contribuir para a minha dissertação.

Por último, resta-me agradecer aos meus colegas de mestrado e trabalho com quem tive oportunidade de trabalhar e partilhar conhecimento.

A todos, um bem-haja!

Índice

1	Introdução	1
1.1	Enquadramento.....	1
1.2	Contexto.....	2
1.2.1	Futebol	2
1.2.2	Tecnologia no Futebol	2
1.3	Problema	3
1.4	Objetivos	4
1.5	Considerações Éticas	5
1.6	Motivação	5
1.7	Metodologia.....	6
1.8	Contributos	7
1.9	Estrutura do Documento	8
2	Estado da Arte	9
2.1	Metodologia de Pesquisa	9
2.2	Business Intelligence.....	10
2.3	Armazém de Dados.....	12
2.3.1	Data Mart.....	13
2.3.2	ETL	13
2.3.3	Modelo Dimensional.....	14
2.4	Ferramentas de ETL	15
2.4.1	Talend	18
2.4.2	Pentaho.....	19
2.4.3	Microsoft SQL Server Integration Services (SSIS).....	20
2.4.4	Análise Comparativa entre Ferramentas ETL	21
2.5	Ferramentas de Análise de Dados	21
2.5.1	PowerBI	23
2.5.2	Qlik Sense	25
2.5.3	Tableau	26
2.5.4	Análise Comparativa entre Ferramentas de Análise de Dados	27
2.6	Sports Analytics	28
2.6.1	Football Analytics	31
2.6.2	Soluções de Sports Analytics.....	33
2.6.3	Websites de Estatísticas e Análises	35
2.7	Casos de Estudo	37
2.7.1	Dashboard para “La Liga”	38
2.7.2	Football Analytics no Liverpool	38
2.7.3	Casos de Sucesso de Saint-Gilloise e Brighton Albion	39
2.8	Conclusões	41

3	Análise e Design	43
3.1	FootyStats	43
3.2	Modelo Relacional	44
3.3	Modelo de Domínio	45
3.4	Engenharia de Requisitos	47
3.4.1	Atores de Sistemas	47
3.4.2	Requisitos	48
3.5	Definição de Indicadores de Performance	53
3.6	Metodologia	55
3.7	Modelo Multidimensional	57
3.8	Arquitetura da Solução	58
3.9	Mapeamento e Transformação de Dados	60
4	Implementação da Solução	67
4.1	Tecnologias Adotadas	67
4.2	Configurações Gerais	68
4.3	Extração	69
4.4	Transformação e Carregamento	81
4.5	Cubo OLAP	94
5	Análise de Resultados e Avaliação da Solução	97
5.1	Análise de Resultados	97
5.1.1	Dashboards	98
5.1.2	Interface Mobile	101
5.2	Avaliação da Solução	102
5.2.1	Metodologia de Avaliação	102
5.2.2	Resultados dos Inquéritos	103
5.2.3	Conclusão	105
6	Conclusão	107
6.1	Objetivos Alcançados	107
6.2	Problemas/Limitações	108
6.3	Trabalho Futuro	109
7	Anexos	117

Lista de Figuras

Figura 1 – Planeamento do projeto	7
Figura 2 – <i>Queries</i> de pesquisa	9
Figura 3 – Esquema circular de dados, informação, conhecimento e sabedoria (Adaptado de [23]).....	10
Figura 4 – Exemplo de arquitetura de BI (Adaptado de [30])	13
Figura 5 – Esquemas Estrela e Floco de Neve [33].....	15
Figura 6 – Quadrante Mágico para ferramentas ETL [37].....	17
Figura 7 – Talend Open Studio [39].....	18
Figura 8 – Exemplo de fluxo de dados configurado em Kettle [40]	19
Figura 9 – Execução de um fluxo de dados configurado via SSIS [41]	20
Figura 10 – Quadrante Mágico para Plataformas de Análise e <i>Business Intelligence</i> [42]	23
Figura 11 – Exemplo de <i>dashboard</i> desenvolvido em PowerBI [43]	24
Figura 12 – Exemplo de <i>dashboard</i> desenvolvido em Qlik Sense [45]	25
Figura 13 – Exemplo de <i>dashboard</i> desenvolvido em Tableau [46]	26
Figura 14 – Categorias de análise de <i>Sports Analytics</i> (Adaptado de [47]).....	28
Figura 15 – Estrutura de <i>Sports Analytics</i> [9], [10], [17]	29
Figura 16 – Proposta de criação de valor no desporto utilizando ferramentas de BI&A por Caya e Bourdon [50]	30
Figura 17 – Modelo de Relacional.....	44
Figura 18 – Modelo de Domínio.....	46
Figura 19 – Diagrama de Casos de Uso	49
Figura 20 – Diagrama de Sequência de Sistema (Consultar Dashboards)	49
Figura 21 – Diagrama de Sequência de Sistema (Filtrar Informação).....	50
Figura 22 – Diagrama de Sequência de Sistema (Validar Informação)	50
Figura 23 – Diagrama de Sequência de Sistema (Gerar Dashboards).....	51
Figura 24 – Diagrama de Sequência de Sistema (Gerir Utilizadores).....	51
Figura 25 – Modelo Dimensional	57
Figura 26 – Arquitetura da Solução.....	59
Figura 27 – Parâmetros do projeto	68
Figura 28 – Conexões do projeto	68
Figura 29 – Variáveis do projeto	69
Figura 30 – Criação da SA.....	70
Figura 31 – Seleção de ligas e extração de primeiro conjunto de tabelas.....	71
Figura 32 – Seleção de ligas e extração de primeiro conjunto de tabelas (Fluxo de Dados).....	71
Figura 33 – Leagues Samples JSON Source (configurações do <i>endpoint</i>).....	72
Figura 34 – Leagues Samples JSON Source (dados)	73
Figura 35 – Seleção de IDs distintos das ligas	74
Figura 36 – Ciclo de extração de dados pendente de LeagueID	74
Figura 37 – Extração de dados pendentes de LeagueID	76
Figura 38 – LeaguePlayers JSON Source (configurações do <i>endpoint</i>)	77

Figura 39 – LeaguePlayers JSON Source (paginação)	77
Figura 40 – Processo de criação da SA	78
Figura 41 – Código de seleção distinta das equipas e jogos completos.....	79
Figura 42 – Extração de dados de jogadores, equipas e continentes	79
Figura 43 – Extração de dados de estádios	79
Figura 44 – Formato JSON do endereço para extração de dados de jogadores_jogos	80
Figura 45 – Extração de diferencial de dados de performance de jogadores em jogos	81
Figura 46 – Criação do AD e respetivas tabelas	82
Figura 47 – Carregamento da DimDate e DimTime_Match	82
Figura 48 – Processo de carregamento da DimDate e DimYear DimTime_Match	82
Figura 49 – Carregamento de restantes Dimensões	83
Figura 50 – Processo de carregamento da DimCountry.....	84
Figura 51 – Processo de carregamento da DimEvent_Match	85
Figura 52 – Processo de carregamento da DimStadium	86
Figura 53 – Processo de carregamento da DimPlayer.....	87
Figura 54 – Processo de carregamento da DimLeague	88
Figura 55 – Processo de carregamento da DimTeam.....	89
Figura 56 – Processo de carregamento da DimMatch	89
Figura 57 – Carregamento da FactLeague_Player.....	90
Figura 58 – Processo de carregamento da FactLeague_Player.....	91
Figura 59 – Carregamento de FactLeague_Team.....	91
Figura 60 – Processo de carregamento de FactLeague_Team.....	92
Figura 61 – Carregamento de FactMatch e FactMatch_Player.....	92
Figura 62 – Processo de carregamento de FactMatch	93
Figura 63 – Processo de carregamento de FactMatch_Player	94
Figura 64 – Cubo OLAP	95
Figura 65 – Hierarquias DimDate	96
Figura 66 – Hierarquia DimMatch_Time	96
Figura 67 – Desempenho das Equipas.....	98
Figura 68 – Desempenho dos Atletas.....	100
Figura 69 – Resumo das Partidas	101
Figura 70 – Interface mobile	102

Lista de Tabelas

Tabela 1 – Comparação entre ferramentas ETL	21
Tabela 2 – Comparação entre ferramentas de análise de dados.....	27
Tabela 3 – Exemplos de dados recolhidos para análises qualitativas e quantitativas no futebol [10]	32
Tabela 4 – Comparação entre soluções de <i>Sports Analytics</i>	35
Tabela 5 – Comparação entre websites de estatísticas e análises no futebol.....	37
Tabela 6 – Dez países europeus mais bem cotados na UEFA [71]	45
Tabela 7 – Atores do Sistema	48
Tabela 8 – Tabela de Mapeamento Lógico de Dados (Excerto)	60
Tabela 9 – Escala de Likert	103
Tabela 10 – Frequência de respostas	104

Acrónimos

API	Interface de Programação de Aplicações
AD	Armazém de Dados
BI	<i>Business Intelligence</i>
BI&A	<i>Business Intelligence & Analytics</i>
B-ON	Biblioteca do Conhecimento Online
BTTS	<i>Both Teams To Score</i>
CIF	Fábrica de Informação Corporativa
DM	<i>Data Mart</i>
DSR	<i>Design Science Research</i>
DQP	<i>Data Quality Platform</i>
ELT	<i>Extraction, Loading and Transformation</i>
ETL	<i>Extraction, Transformation and Loading</i>
FC ou CF	<i>Football Club</i>
FIFA	Federação Internacional de Futebol
FM	<i>Football Manager</i>
FT	<i>Full Time</i>
FURPS+	<i>Functionality, Usability, Reliability, Performance, Supportability</i>
GPS	Sistema de Posicionamento Global
HT	<i>Half Time</i>
ID	Identificador

IT ou TI	Tecnologias de Informação
IoT	<i>Internet of Things</i>
ISO	Organização Internacional de Normalização
JSON	<i>JavaScript Object Notation</i>
KPI	<i>Key Performance Indicator</i>
NoSQL	<i>Not Only Structured Query Language</i>
OLAP	<i>Online Analytical Processing</i>
PPG	<i>Points Per Game</i>
REST	<i>Representational State Transfer</i>
SA	<i>Staging Area</i>
SCD	<i>Slow Changing Dimension</i>
SGBD	Sistema de Gestão de Base de Dados
SSIS	<i>SQL Server Integration Services</i>
SQL	<i>Structured Query Language</i>
SSC	<i>Società Sportiva Calcio</i>
UEFA	União das Associações Europeias de Futebol
VAR	Árbitro Assistente de Vídeo
UML	<i>Unified Modeling Language</i>

1 Introdução

Neste Capítulo é descrito um enquadramento sobre o projeto, identificado o problema e objetivos do mesmo. Adicionalmente, são discutidas as considerações éticas relevantes, mencionada a motivação que levou à escolha do tema, bem como, a metodologia usada e os seus contributos. Por fim, é apresentada a estrutura do presente documento.

1.1 Enquadramento

O desporto desempenha um papel fundamental em toda a sociedade, especialmente o futebol uma vez que se trata de um dos desportos mais populares do mundo, no qual os dados estão a ganhar cada vez mais uma maior preponderância [1], [2].

No presente século, existe uma batalha analítica onde os dados e a sua apreciação podem determinar se uma equipa terá ou não sucesso desportivo. Clubes de todo o mundo têm vindo investir progressivamente neste tipo de técnicas, comprovando o seu valor [3].

A análise de dados tem vindo a ser aplicada em diversos campos desde a monitorização de atletas, *scouting*, prevenção de lesões, previsões de comportamentos em campo, entre outros. Casos de sucesso como os de Liverpool FC [4], [5], Brighton & Hove Albion FC e Royal Union Saint-Gilloise [3] são evidências de que estes recursos acarretam valor para o futebol.

No âmbito de análise de dados, o *Business Intelligence* (BI) é um dos variados recursos computacionais que tem conquistado o seu espaço no futebol, sobretudo devido ao grande volume de dados que consegue armazenar e analisar [6].

1.2 Contexto

A presente Secção tem por objetivo aprofundar o contexto na qual a dissertação será alvo de estudo. Nesta será apresentado o conceito de futebol e como a tecnologia tem emergido nesse mesmo setor. Compreender estes conceitos é fundamental, uma vez que toda a composição do documento será baseada nos mesmos.

1.2.1 Futebol

O futebol, um desporto apaixonante e de grande impacto global, é praticado por duas equipas de onze jogadores em campo retangular delimitado por balizas. Com um guarda-redes e dez jogadores de campo, o objetivo é simples: marcar golo na baliza adversária. A partida é dirigida por um árbitro principal, auxiliado por dois fiscais de linha e um quarto árbitro, além de árbitros assistente de vídeo (VAR) em certas competições [7]. A Federação Internacional de Futebol (FIFA) é a entidade máxima que regulamenta e gere o futebol mundial, organizando eventos como o Campeonato do Mundo de Clubes e o Campeonato Mundial de Futebol [8]. A União das Associações Europeias de Futebol (UEFA), por sua vez, lidera o futebol no continente europeu, organizando competições como a Liga dos Campeões, a Liga Europa e a Liga Conferência, impulsionando não apenas o desporto, mas também o turismo e a economia dos países envolvidos. O sucesso de um clube nessas competições não só traz benefícios financeiros como também eleva o coeficiente do país na UEFA, garantindo mais vagas para futuras participações [3]. Este círculo virtuoso é crucial para o desenvolvimento e a competitividade do futebol europeu. A FIFA considera que rotular o futebol apenas como desporto é demasiado redutor, já que representa um enorme valor para toda a humanidade enquanto sinónimo de inclusão e de transmissão de valores como disciplina, resiliência, integridade, *Fair Play*, entre outros [9], [10].

1.2.2 Tecnologia no Futebol

O futebol, assim como qualquer outro desporto, é imprevisível, com a sorte ocasionalmente desempenhando um papel decisivo, o que contribui para o seu encanto universalmente admirado [11]. Neste contexto, é da responsabilidade do treinador, no meio da aleatoriedade do jogo, avaliar a situação presente e determinar a sequência de eventos que conduzirá ao desfecho desejado. Para tal, o acesso a dados históricos individuais e coletivos das equipas é

crucial, facilitando todo o processo decisório [12]. O interesse pela análise no futebol tem crescido exponencialmente entre investigadores e órgãos reguladores, como a FIFA [9], [10], [13], que prevê que em 2026 mais da metade da população mundial estará de alguma forma envolvida com o desporto [14]. O sucesso de desportos como basquetebol e beisebol na análise de dados serve como exemplo para o futebol [13], onde técnicas estatísticas, probabilísticas e de *machine learning* têm sido aplicadas para apoiar decisões táticas e estratégicas dos treinadores [9], [10]. A evolução tecnológica, com destaque para sistemas de vídeo e dispositivos de monitorização biométrica como GPS e sensores, tem desempenhado um papel crucial na análise de desempenho dos jogadores e na melhoria das decisões arbitrais [11], [12]. A introdução de tecnologias, como a tecnologia de linha de golo e o VAR, evidenciam o avanço tecnológico no futebol e a sua aceitação progressiva pelas principais instituições desportivas [12]. As tecnologias de localização, como o GPS, têm sido amplamente adotadas, fornecendo dados precisos sobre a performance dos atletas, que combinados com técnicas de *Business Intelligence* permitem uma análise mais profunda e significativa [15]. A UEFA e a FIFA têm investido em plataformas de inovação para promover o desenvolvimento de novas tecnologias no futebol, visando não só melhorar o desempenho desportivo, mas também aumentar o envolvimento dos adeptos [12]. Este crescente enfoque na análise de dados e tecnologia reflete uma tendência global na indústria do futebol, em busca de um jogo mais justo, eficiente e emocionante.

1.3 Problema

Existência de uma nova era no desporto caracterizada pela enorme quantidade de dados descentralizados, onde é impensável o ser humano realizar análises desportivas sem recurso às Tecnologias de Informação (TI) [6], [16].

Muito clubes desportivos operam sobre uma gestão muito tradicional, sobretudo no departamento de desporto, no qual raramente está presente um colaborador da área de TI. Além disso, apresentam resistência no que toca à partilha de dados [6].

Adicionalmente, observa-se uma escassez de referências na literatura acerca da adoção de tecnologias de armazenamento e ferramentas de processamento analítico de dados, visando a toma de melhores decisões desportivas nos clubes. O cenário mais comum passa pela existência de bases de dados relacionais que reúnem diversas estatísticas de ligas, equipas e jogadores.

No entanto, estas apresentam diversas fragilidades tanto ao nível da esquematização de dados, como também ao nível das análises que permitem.

É essencial traduzir esses dados de maneira que todos possam aceder às informações e obter conhecimento a partir deles. Por outras palavras, os dados devem responder de maneira objetiva à pergunta original. Adicionalmente, é crucial que as análises sejam conduzidas utilizando uma linguagem do dia a dia, de modo que todos na organização do futebol possam entendê-las facilmente [17].

Para além do já mencionado, uma das limitações mencionadas na literatura [18], é a escalabilidade de ferramentas de processamento analítico na análise desportiva para múltiplas competições, devido ao tempo na obtenção de dados, o que será um dos tópicos de estudo nesta dissertação.

1.4 Objetivos

O desporto gera enormes quantidades de dados, principalmente representados sob a forma de base de dados relacionais. No entanto, carecem de funcionalidades que permitam análises mais eficazes e um entendimento global por todo o mundo no futebol. Neste sentido, a presente dissertação tem como objetivo global, dotar todos os profissionais de futebol, sejam eles treinadores, olheiros², diretores desportivos, jogadores, etc., de uma ferramenta capaz de fornecer informação determinante para a tomada de melhores decisões desportivas, baseada em dados objetivos e detalhados de atletas, equipas e competições.

De forma a atingir-se o objetivo global mencionado anteriormente, é necessário estabelecer metas concretas. Estas são constituídas pelas seguintes:

- Analisar diferentes métodos de recolha de dados e fontes de dados;
- Conceber a arquitetura do armazém de dados;
- Elaborar e executar processos de extração, transformação e carregamento (ETL) relevantes para o carregamento dos dados no armazém de dados;

² Olheiro - Profissional responsável pela análise e observação, cujo principal objetivo é obtenção do máximo de informação possível tanto a nível individual como coletivo. Termo muitas vezes utilizado na identificação de profissionais que procuram potenciais jovens atletas promissores.

- Conceber cubos de processamento analítico online (OLAP) para a análise de jogadores e equipas;
- Produzir métricas adicionais através dos dados carregados nos cubos;
- Desenhar e estruturar *dashboards* personalizáveis, intuitivos e eficazes;
- Analisar os resultados obtidos;
- Documentar todo o processo envolvido.

1.5 Considerações Éticas

O desenvolvimento de soluções destinadas a lidar e apresentar dados no contexto do futebol implica considerações éticas profundas, com o objetivo de assegurar a integridade e equidade, tanto na manipulação dos dados envolvidos, como na utilização dessas tecnologias. A extração e tratamento dos dados devem respeitar a privacidade dos indivíduos envolvidos, seguindo normas éticas e legais para a recolha e tratamento de dados pessoais, garantindo o consentimento informado e a proteção da confidencialidade.

Além disso, é crucial ponderar sobre a transparência e a equidade dos dados apresentados. A clareza na comunicação e a acessibilidade da informação a todas as partes interessadas são fundamentais para evitar interpretações tendenciosas, promovendo um contexto de análise justo no âmbito desportivo. Por fim, a responsabilidade social também deve ser considerada, assegurando que a solução contribua para o bem-estar geral da comunidade desportiva e promova os valores inerentes ao futebol.

1.6 Motivação

O futebol é um desporto muito famoso, o qual é acompanhado e praticado por biliões de pessoas globalmente. A análise de dados desempenha um papel vital no universo do futebol, revolucionando a forma como as pessoas olham para os jogos, na organização das equipas, na seleção dos jogadores e em diversos outros aspetos. Esta prática, permite dotar os clubes de uma melhor compreensão do seu jogo e, conseqüentemente, obter resultados superiores [1].

A escolha do tema prende-se essencialmente no fascínio que o autor possui na área do futebol conjuntamente com a análise de dados, desde a sua infância. Adicionalmente, a oportunidade

de acrescentar valor para o futebol, bem como, aprofundar o conhecimento nesse domínio foi decisivo para a escolha do referido tópico.

A solução a desenvolver torna-se atrativa, não apenas por se enquadrar num tema com enorme potencial de crescimento, mas também por oferecer oportunidades de aperfeiçoar o uso de ferramentas e aprofundar o conhecimento adquirido sobre *Business Intelligence*, o que é altamente valorizado no contexto profissional.

1.7 Metodologia

O presente trabalho será desenvolvido usando a metodologia *Design Science Research* (DSR). Esta define-se por ser um paradigma que visa aumentar o conhecimento tecnológico e científico de um domínio específico, mediante a conceção de artefactos inovadores que abordam problemas particulares. A DSR é composta pelas seis fases seguintes [19]:

- I. Identificação do problema e motivação, os quais foram descritos nas Secções 1.3 e 1.6, respetivamente;
- II. Definição dos objetivos da solução, que foram enumerados na Secção 1.4;
- III. Design e desenvolvimento, no qual o design é abordado no Capítulo 3, onde são delineados os requisitos funcionais e não funcionais, os atores e a arquitetura do sistema. O desenvolvimento é endereçado no Capítulo 4, através da documentação de todo o processo de implementação;
- IV. Demonstração, a qual é alvo de foco na Secção 5.1, através da apresentação das funcionalidades desenvolvidas;
- V. Avaliação, realizada através da apreciação por parte de uma equipa especializada no domínio do futebol sob a forma de questionário, o qual é exibido na Secção 5.2;
- VI. Comunicação, que é efetuada através do presente documento.

Esta metodologia é indicada para este tipo de contexto devido à possibilidade de melhorar os artefactos anteriormente produzidos através de um processo de feedback e construção, de forma gradual e dinâmica [20].

Na Figura 1, é exposto o planeamento abrangente do projeto, incluindo a duração de cada fase, coincidindo com as etapas acima mencionadas.

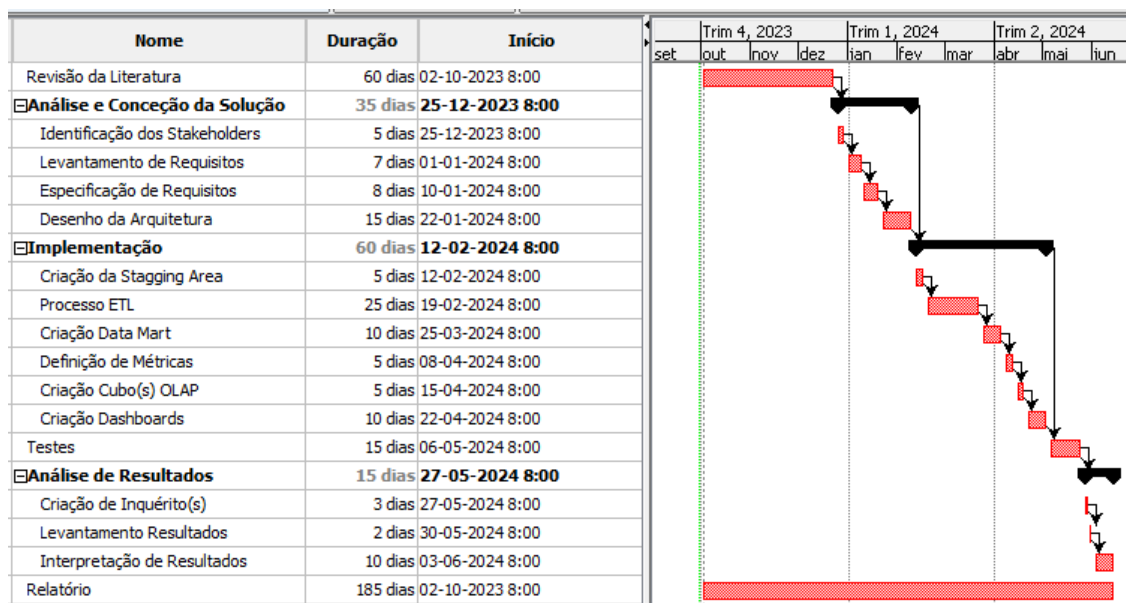


Figura 1 – Planeamento do projeto

1.8 Contributos

Conforme mencionado anteriormente, este projeto visa desenvolver um sistema de *Business Intelligence*, uma abordagem que tem sido pouco explorada na literatura, visando aprimorar a fundamentação e eficiência nas decisões tomadas pelos profissionais de futebol. Esta solução permitirá o acesso a informação e proporcionará uma experiência personalizável, intuitiva e eficaz, através de dashboards.

Os *dashboards* reunirão um conjunto de *Key Performance Indicators* (KPI³), cuidadosamente organizados e apresentados sobre a forma de um ecrã, permitindo que um clube tenha acesso a informação dinâmica e atempada, facilitando e acelerando a compreensão por parte de todos profissionais de futebol. Contrariamente aos relatórios, que oferecem representações estáticas de informação sobre um período específico, exigindo a sua recriação em caso de alterações [21].

Uma aplicação com estas características pode trazer inúmeros benefícios para a comunidade do futebol, uma vez que irá proporcionar *insights* extremamente valiosos como, por exemplo,

³ KPI – *Key Performance Indicator*, designação que em português se traduz como indicador chave de desempenho, constitui uma representação das métricas por meio das quais as organizações avaliam sua eficiência.

apoiar nas escolhas táticas de um treinador e auxiliar na identificação do atleta mais adequado para desempenhar a função pretendida de forma mais eficiente [21].

1.9 Estrutura do Documento

O presente documento é constituído por seis Capítulos.

O primeiro Capítulo apresenta um enquadramento e contexto sobre o projeto, definindo o problema, e os objetivos pretendidos. São também mencionadas as considerações éticas, a motivação que levou à escolha do tema, metodologia usada durante a realização da dissertação e os seus contributos.

No segundo Capítulo, o Estado da Arte, é realizada uma revisão literária sobre os conceitos de *Business Intelligence*, Armazéns de Dados, ETL e Modelo Dimensional. Ferramentas de ETL e de Análise de Dados são, também, apresentadas. Adicionalmente, é abordado o conceito de *Sports Analytics*, identificadas soluções e websites incorporadas nesse domínio e efetuadas comparações entre elas. No final, são exibidos casos de estudo e apuradas conclusões.

No terceiro Capítulo, análise e design, são definidos os requisitos funcionais e não funcionais e apresentada a arquitetura da solução.

No quarto Capítulo, implementação da solução, são enunciadas as ferramentas utilizadas durante a realização do projeto, bem como descrito o processo de ETL e de geração de cubos OLAP, com base na arquitetura apresentada. Tendo em conta as constatações identificadas nos pontos anteriores, são também identificadas as decisões tomadas e os motivos a elas associados.

No quinto Capítulo, análise de resultados e avaliação da solução, são apresentados os resultados obtidos nos *dashboards* desenvolvidos e discutidos os resultados alcançados através deles. Além disso, são também apresentados os métodos de avaliação da solução.

Por fim, o Capítulo seis, no qual são descritas as conclusões do projeto desenvolvido, enunciadas as limitações obtidas e referenciados pontos de melhoria para trabalho futuro.

2 Estado da Arte

Neste Capítulo serão aprofundados os termos de *Business Intelligence* e armazéns de dados, bem como outros conceitos subjacentes. Serão apresentadas e comparadas algumas ferramentas de ETL e de análise de dados. Explorar-se-á o conceito de *Sports Analytics*, expostas diversas soluções tecnológicas existentes e efetuada uma comparação entre as mesmas. Casos de estudo são referidos e, por último, são retiradas conclusões sobre a revisão da literatura realizada.

2.1 Metodologia de Pesquisa

De forma a reunir toda a informação pertinente para a revisão da literatura, foi utilizada a biblioteca do conhecimento online (b-on) conjuntamente com o motor de busca Google Académico. Durante a pesquisa foram utilizados os operadores lógicos “and” e “or” com a finalidade de constituir uma *query* suficientemente robusta capaz de devolver informação objetiva. Após diversas tentativas, a consulta final ficou definida pelas duas *queries* seguintes (Figura 2):

```
((Football OR SOCCER or Futebol) & ((DATA warehouse) OR (Armazéns de Dados)))  
OR ((DATA warehouse in football) OR (Armazéns de dados no futebol))  
  
((Football OR SOCCER or Futebol) & (Business Intelligence))  
OR ((Football OR SOCCER or Futebol) & ((Sistemas de processamento analítico) OR (Analytic Process SYSTEM) OR (OLAP)))
```

Figura 2 – *Queries* de pesquisa

Para refinar ainda mais a pesquisa foram estabelecidos os critérios de (1) apenas considerar dissertações, artigos, revistas e conferências disponíveis para consulta e transferência, (2) 10 anos o tempo máximo de publicação dos documentos, isto é, até no máximo 2013 e (3)

documentos redigidos apenas nas línguas portuguesa, inglesa, espanhola e italiana. Após a aplicação destes critérios, foram considerados quarenta artigos.

2.2 Business Intelligence

A informação é a peça fulcral para qualquer organização, sendo a principal responsável por transformar qualquer negócio comum, num negócio inteligente. Esta, uma vez que advém da compilação e estruturação de dados em diversos formatos, deve ser gerida, organizada e combinada de forma cuidadosa para que seja passível de se analisar e obter conhecimento [22]. A sabedoria, por sua vez, é alcançada através do conhecimento, facilitando todo o processo de tomada de decisão (Figura 3). Assim sendo, quanto melhor for a qualidade da informação, mais consistente será o conhecimento, mais prudente será a sabedoria e mais bem fundamentadas serão as decisões tomadas [23], [24].

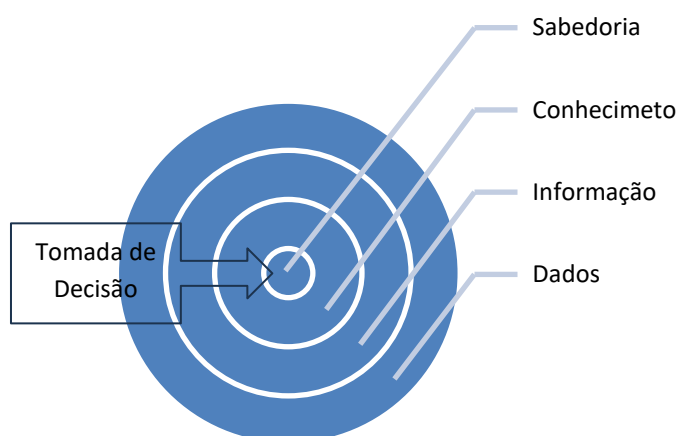


Figura 3 – Esquema circular de dados, informação, conhecimento e sabedoria (Adaptado de [23])

O termo *Business Intelligence* (BI) é um conceito muito presente na atualidade que teve a sua origem datada na década de 1890 [25]. No entanto, na literatura não existe consensualidade quanto à sua definição.

O BI, também conhecido por *Business Analytics*, pode ser definido como um conjunto de ferramentas capazes de transformar os dados acessíveis em informação relevante para a tomada de decisões inteligentes para o negócio [17], [18], [20], [26], [27]. Gonzalez Ardevínez em [23], por sua vez, indica que o BI é um “processo de análise do desempenho e da eficiência” de uma organização sobre uma perspetiva histórica. Enquanto que Ratten V e Dickson G [27], definem BI como a combinação entre produtos e processos inteligentes capazes de trabalhar

com conhecimento, visando decisões mais sustentadas. Posto isto, pode-se afirmar que, apesar de não existir definição consensual quanto a BI, o propósito identificado na literatura é comum, sendo este, auxiliar na tomada de decisão de uma organização [11]. Neste contexto, a definição de Gonzalez Ardevínez [23] é adotada nesta dissertação, uma vez que se procura promover a eficiência de um clube através do desempenho dos seus atletas ao longo do tempo, o que está alinhado com os objetivos delineados.

O BI é caracterizado pelas seguintes etapas [2], [28]:

1. Recolha de dados: Todos os dados relacionados ao negócio são considerados com o objetivo de identificar gargalos, potenciais oportunidades de mercado e estratégias para melhorar o desempenho, entre outros aspetos;
2. Organização e análise: Os dados são organizados num armazém de dados e exibidos sobre uma forma de *dashboards* e, ou relatórios, facilitando a consulta e análises a serem realizadas por indivíduos encarregados de tomar decisões;
3. Ação e monitorização: As ações são tomadas com base em consultas de informação previamente realizadas. Enquanto que, a monitorização dos resultados é realizada através de métricas de desempenho estabelecidas.

Para além do BI agilizar a tomada de decisão de uma organização, o que lhe favorece competitivamente no mercado onde se insere, este permite também integrar dados históricos heterogéneos, melhorar qualitativamente os dados, consultar informação constantemente atualizada, sobre uma perspetiva abrangente ou detalhada, apurar e acompanhar a evolução dos KPI's fundamentais para o negócio, entre outros [11]. No entanto, projetos de *Business Intelligence* são morosos, árduos de implementar e possuem um elevado custo e risco, tornando impeditiva a sua implementação para algumas empresas. Considerar-se implementações desta índole como meros projetos de TI, em vez de uma estratégia, visão, arquitetura dinâmica e escalável ao longo do tempo, tendo como finalidade o alinhamento entre objetivos operacionais e os estratégicos, é apontada como a principal razão para o insucesso [29].

No futebol, a integração do BI tem sido um processo gradual, transformando o método de como as análises desportivas são realizadas e como, no desporto, a tomada de decisão é consumada. Esta tecnologia tem vindo a permitir recompilar inúmeras informações de equipas, partidas e

jogadores, proporcionando uma imagem clara e objetiva do que ocorre dentro das quatro linhas e de como melhorar o desempenho individual e coletivo, algo que antes era realizado de forma empírica [18]. Segundo Fontanive L [9], a capacidade de um atleta está relacionada com dois elementos, dados e desempenho, e o BI apresenta um conjunto ideal de ferramentas capazes de ajudar o jogador a alcançar seus objetivos individuais, através da exposição de um panorama completo do seu rendimento.

2.3 Armazém de Dados

Um armazém de dados (AD) consiste num repositório central de dados multidimensional com a finalidade de retirar-se análises que apoiem as decisões empresariais [3], [30]. Estas podem ser executadas via ferramentas de visualização de dados ou através de técnicas de mineração de dados [20]. Os armazéns de dados são constituídos por dados históricos provenientes de fontes de dados heterogéneas e heteromorfas [30]. Para além disso, um AD é orientado ao assunto, sendo os dados organizados em torno dos principais assuntos e, não volátil, apenas permitindo operações de carregamento e leitura de dados [30]. Em [21] acrescenta-se que um AD possui duas vertentes, uma técnica, responsável por extrair, limpar e organizar dados vindouros de sistemas operacionais para serem carregados em estruturas multidimensionais, e outra empresarial, incumbida de tornar as consultas e análises de informação possíveis para a toma de melhores decisões.

Ao nível da sua arquitetura global, um AD é composto por uma *back room* ou *backend*, onde é realizado todo o processo de extração, transformação e carregamento (ETL) de dados, e uma *front room* ou *frontend*, sendo esta a interface através da qual o utilizador interage com o AD e toma as decisões [21]. No seu detalhe, um armazém de dados é composto por [30]:

- Fontes de dados: Um AD é constituído por diversas fontes de dados, são exemplos de bases de dados relacionais, ficheiros CSV, JSON, ente outros;
- *Staging area*: Área onde os dados são submetidos ao processo de ETL. Estes sofrem diversas transformações como limpeza, concatenações, cálculos, de forma a preparar-se os dados para as análises;
- Armazenamento: Estrutura multidimensional que armazena os dados do negócio. Um AD é constituído por dois ou mais *Data Marts* (DM);

- *Analytics*: Componente através da qual é possível efetuar-se consultas, análises e, ou mineração de dados.

Na Figura 4, é possível verificar-se em detalhe o diagrama da arquitetura mencionada.

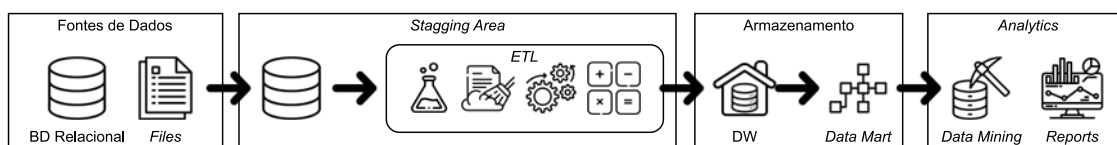


Figura 4 – Exemplo de arquitetura de BI (Adaptado de [30])

Bill Inmon e Ralph Kimball, pioneiros no campo dos armazéns de dados, estabeleceram duas arquiteturas para a execução de projetos de BI [3], [21], [23]. Inmon introduziu a arquitetura de fábrica de informação corporativa (CIF), caracterizada por ser uma abordagem *Top-down*, onde os dados são extraídos, transformados e carregados num AD antes de serem distribuídos por diversos *Data Marts*. Kimball, por sua vez, propôs uma abordagem *Bottom-up*, privilegiando a criação de *Data Marts* de cada área de negócio, sendo o AD obtido através da combinação desses *Data Marts*. Esta arquitetura, para além de constituir-se uma metodologia ágil, altamente valorizada no mercado, simplifica um processo de implementação de inteligência de negócio [31].

2.3.1 Data Mart

Define-se um *Data Mart* (DM), uma estrutura multidimensional integrada num AD, capaz de responder às necessidades de negócio de um grupo de indivíduos, departamento numa organização, ou clube de futebol. Face aos armazéns de dados, estes tendem a ser mais rápidos tanto ao nível da sua implementação, como ao nível das consultas devolvidas [3], [21]. Em termos de sistema, os DMs podem esquematizar-se sobre um formato em estrela, constelação ou floco de neve [23].

2.3.2 ETL

O ETL (correspondente às iniciais em inglês de *Extraction, Transformation e Loading*) constitui um processo trifásico determinado pelas operações de extração “dos dados sistemas fonte” [20], transformação, na qual são aplicadas alterações à forma dos dados para um formato

multidimensional e, carregamento de dados num AD, retratando o procedimento mais moroso no contexto do BI [18], [20].

A qualidade dos dados desempenha um papel fundamental na utilidade das análises, o que justifica a necessidade de um investimento significativo de tempo na transformação dos dados, sendo uma das razões fundamentais para a morosidade do processo de ETL. Para além disso, dado que se trata de um processo dinâmico sujeito a mudanças nas fontes de dados, é imperativo desenhar-se uma arquitetura suficientemente robusta e escalável para que a mesma seja mantida ao longo do tempo [18], [20].

Conforme mencionado no Subcapítulo anterior, os ADs são caracterizados por possuir um segmento técnico e outro empresarial, sendo o ETL incluído na componente técnica [21].

Uma alternativa ao ETL é o ELT, onde a extração é realizada de forma idêntica ao ETL, no entanto, os dados são diretamente carregados no AD e, posteriormente, executadas as transformações de dados. Estes processos são muitas vezes implementados em sistemas cujas necessidades se regem por uma maior rapidez e um maior fluxo contínuo de dados [21].

2.3.3 Modelo Dimensional

O modelo dimensional, um conceito estabelecido por Kimball, define-se como uma estrutura de dados simples e desnormalizada na qual se baseia a construção de um *Data Mart*, facilitando as análises a realizar-se sobre uma perspetiva histórica. Adicionalmente, este modelo possibilita a criação de cubos de processamento analítico online (OLAP), os quais possibilitam a realização de consultas com um desempenho consideravelmente superior em comparação com um modelo relacional [23]. Nesta esquematização, os dados são organizados em tabelas de factos e tabelas de dimensão dispostas num formato estrela, constelação ou floco de neve [18].

2.3.3.1 Tabelas de Factos e de Dimensão

As tabelas de dimensão são caracterizadas por serem tabelas compostas por diversos dados descritivos que ocupam pouco espaço comparativamente às tabelas de factos, providenciando contexto para as análises. Por sua vez, as tabelas de factos são compostas por dados numéricos e outros factos que se pretendem observar/analisar [23]. Os factos podem ser aditivos, podendo ser agregados sobre todas as dimensões, semi-aditivos, apenas agregados sobre algumas dimensões e não aditivos, não podendo ser agregados sobre qualquer dimensão [32].

2.3.3.2 Modelos Estrela, Constelação, Floco de Neve

O modelo em estrela, é identificado pelo relacionamento das tabelas de dimensão com uma tabela de factos central, formando uma estrela. O modelo constelação é composto por vários modelos estrela interligada, nos quais se verifica a presença de mais de uma tabela de factos. Por último, o modelo em floco de neve é caracterizado por ser muito idêntico ao formato estrela, no entanto, apresenta alguma normalização nas tabelas de dimensão, isto é, apresenta ligações entre pelo menos duas dimensões [18]. Na Figura 5, é possível observar-se os modelos estrela e floco de neve.

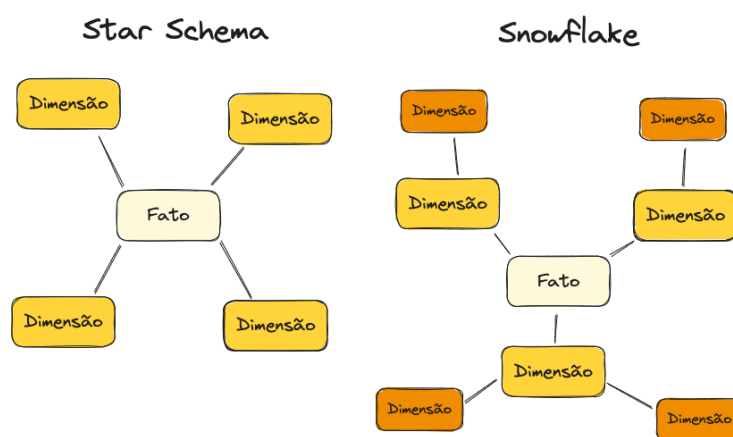


Figura 5 – Esquemas Estrela e Floco de Neve [33]

2.4 Ferramentas de ETL

As ferramentas de ETL possibilitam executar todo o processo de extração, transformação e carregamentos de dados, proporcionando às organizações dados relevantes capazes de auxiliarem na tomada de decisão e de serem utilizados por outros sistemas. Existem sete categorias pelas quais é possível identificar este tipo de ferramenta, podendo, em alguns casos, pertencerem a mais do que uma. Estas categorias incluem, processamento em lote, baseados em código ou motor, baseados em nuvem, *open source*, baseadas em interface, *real-time* e baseadas em *Not Only Structured Query Language (NoSQL)* [34].

De forma a comparar as ferramentas, é utilizado o Quadrante Mágico da Gartner, que através de um conjunto de critérios, avalia se uma solução pertence a um dos quatro quadrantes seguintes [35]:

- Líderes – Ferramentas que no momento atual desempenham extraordinariamente o seu propósito, encontrando-se igualmente bem preparadas para enfrentar os desafios futuros;
- Desafiantes – Soluções que no presente cumprem excepcionalmente com o seu propósito ou têm potencial para conquistar um grande segmento de mercado, contudo, carecem de uma compreensão clara da direção a ser seguida no mercado;
- Visionários – Demonstram compreensão acerca da dinâmica do mercado ou apresentam uma perspectiva capaz de influenciar as leis do mercado, no entanto, não concretizam eficazmente o seu propósito;
- Concorrentes de Nicho – Operam sobre pequenos segmentos de mercado.

No que diz respeito às ferramentas de ETL, a Gartner realiza uma avaliação considerando vários critérios. No entanto, neste projeto, serão considerados apenas os seguintes oito critérios [36]:

- Integração de dados operacionais – Viabilizar a integração de dados provenientes de sistemas operacionais/transacionais, assim como a sincronização de dados essenciais aos processos cruciais de uma organização;
- Virtualização dos dados – Possibilitar a consulta de dados heterógenos previamente integrados, por meio de adaptadores de fontes de dados, um repositório de metadados, ou motor de consulta distribuído;
- Integração de dados em fluxo contínuo – Capacidade de processar dados em movimento e providenciar esses mesmos dados para consumo posterior;
- Serviços API – Capacidade de integrar e disponibilizar dados para utilização através de uma Interface de Programação de Aplicações (API);
- Transformação de dados complexos – Facilidade no desenvolvimento de operações complexas de modelação e processamento de dados, permitindo também a sua reutilização;
- Integração de dados melhorada – Inclusão de diversas estratégias para melhorar a integração de dados, seja via metadados ou algoritmos de *machine learning*;
- Preparação dos dados – Nível de adaptação da ferramenta de forma a possibilitar a sua utilização por utilizadores com diversos níveis de conhecimento;

- Portabilidade da integração – Possibilidade de movimentar o fluxo de dados desenvolvido entre infraestruturas, preservando os níveis de segurança e confidencialidade.

Através da Figura 6, é possível observar-se o posicionamento de mercado das ferramentas de ETL sobre o Quadrante Mágico do Gartner no passado ano de 2022.



Figura 6 – Quadrante Mágico para ferramentas ETL [37]

As ferramentas de ETL selecionadas para investigação na revisão literária foram Talend, Pentaho e Microsoft SQL Server *Integration Services* (SSIS). A decisão foi fundamentada com base no Quadrante Mágico da Gartner apresentado, onde duas são destacadas como líderes, enquanto a Hitachi Vantara (Pentaho) é apresentada como uma concorrente de nicho.

2.4.1 Talend

Talend é uma organização cujo seu negócio é providenciar soluções e serviços especializados em integração, gestão e armazenamento de diversos dados em *cloud*, permitindo que as empresas efetuem decisões em tempo real, mantendo elevados padrões de qualidade [38].

Talend Open Studio, objeto de análise neste projeto, constitui uma dessas soluções. Esta define-se por ser uma ferramenta ETL *open source* capaz de integrar dados através de uma interface de fácil utilização, proporcionando, desse modo, uma redução significativa no tempo necessário para a sua implementação [38]. Na Figura 7, observa-se um fluxo de dados desenvolvido utilizando o Talend Open Studio.

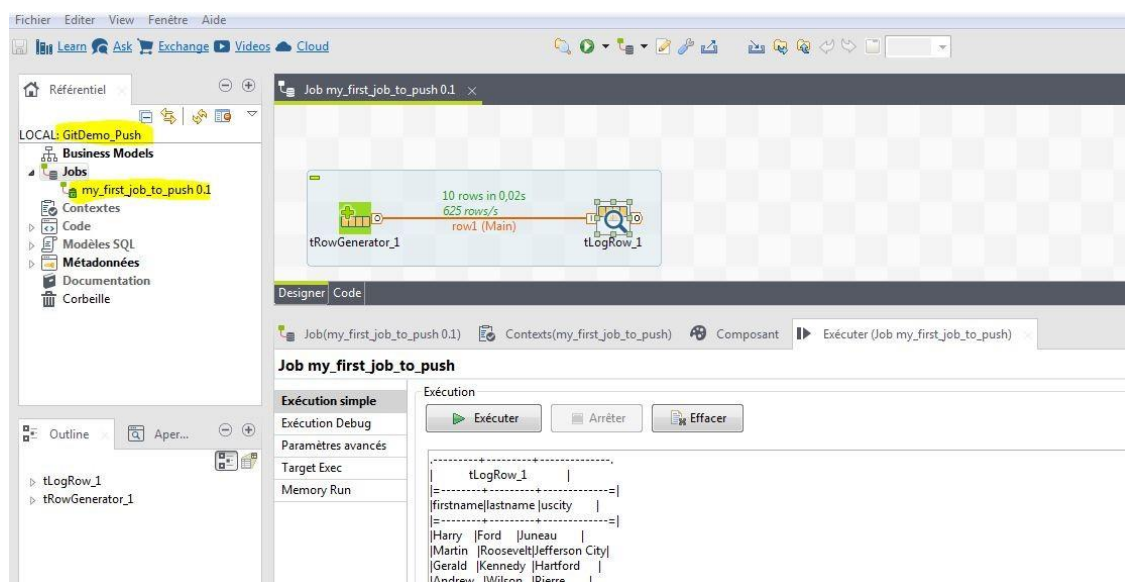


Figura 7 – Talend Open Studio [39]

Gartner destaca o compromisso na melhoria de qualidade dos dados durante a criação dos fluxos de trabalho, o aperfeiçoamento das capacidades de autosserviço na preparação e engenharia dos dados, e o suporte facultado em cenários de replicação de dados como sendo os principais destaques da Talend, na avaliação realizada no ano de 2022. Por outro lado, realça a necessidade de melhorias na observabilidade dos dados, suporte prestado aos clientes e preços praticados [36].

2.4.2 Pentaho

Pentaho, vulgarmente conhecida por Kettle, é uma ferramenta ETL detida pela empresa Hitachi Vantara capaz de integrar, limpar e armazenar dados adotando um formato uniforme passível de ser consumido por utilizadores finais e tecnologias *Internet of Things* (IoT) [40].

Outras soluções oferecidas pela Pentaho incluem o Pentaho BI Server, o Pentaho Meta Editor, o Pentaho Schema Workbench, o Pentaho Business Analytics, o Pentaho Reporting Designer, o Pentaho Dashboards Designer e o Pentaho Data Mining [40].

Na figura 8 verifica-se a execução de um fluxo de trabalho desenvolvido através do Kettle.

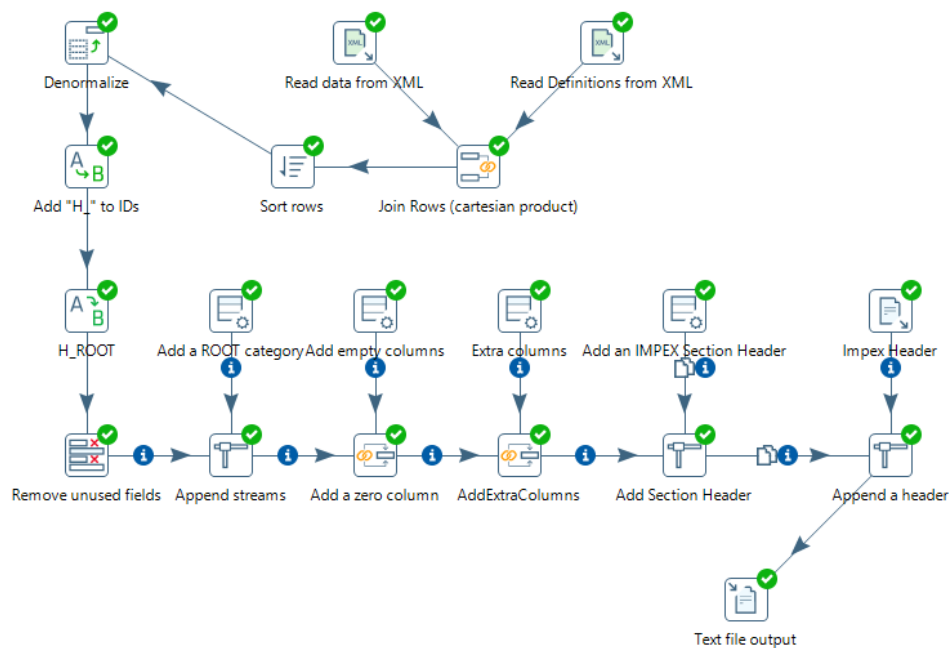


Figura 8 – Exemplo de fluxo de dados configurado em Kettle [40]

Na análise efetuada no ano de 2022, Gartner sublinha a aprimorada competência das soluções de Hitachi Vantara na gestão de dados, integração com dados provenientes da IoT e suporte facultado para integração de dados em segmentos específicos. No entanto, são evidenciadas algumas limitações, tais como dependência exclusiva do Pentaho para a tração de mercado, suporte embrionário para componentes distintos da estrutura de dados e restrições nas implementações nativas da nuvem [36].

2.4.3 Microsoft SQL Server Integration Services (SSIS)

O SQL Server Integration Services (SSIS), uma ferramenta *on-premise* proprietária da Microsoft, destaca-se pela sua capacidade de abordar desafios complexos relacionados com extração, transformação e carregamento de dados num contexto empresarial. Adicionalmente, a ferramenta é utilizada para automatizar a administração de bases de dados SQL Server e as atualizações de cubos multidimensionais [36]. Na figura 9 observa-se a execução de um fluxo de trabalho desenvolvido através do SSIS.

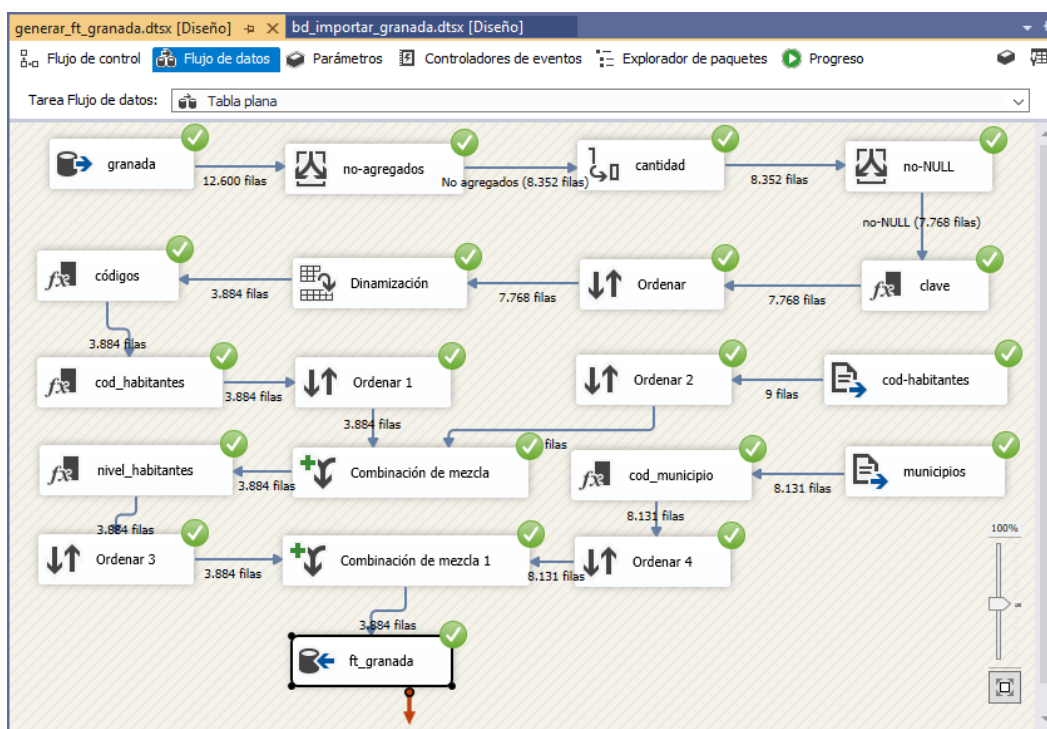


Figura 9 – Execução de um fluxo de dados configurado via SSIS [41]

Através da avaliação realizada em 2022, Gartner realçou a notável estratégia do produto, focada em casos de uso empresariais e departamentais com a maior força do SSIS. Contudo, foram identificadas fragilidades significativas, tanto ao nível da depuração e relatório de erros, como também em relação às limitações na integração de arquiteturas híbridas [36].

2.4.4 Análise Comparativa entre Ferramentas ETL

A Tabela 1 destaca a análise comparativa das ferramentas de ETL selecionadas, com base nos critérios da Gartner escolhidos para avaliação no contexto deste projeto.

Tabela 1 – Comparação entre ferramentas ETL

Critérios	Talend	Pentaho (Kettle)	Microsoft SSIS
Integração de dados operacionais	☒	☒	☒
Virtualização dos dados	☒	☒	☒
Integração de dados em fluxo contínuo	☒	☒	☒
Serviços API	☒	☒	☒
Transformação de dados complexos	☒	☒	☒
Integração de dados melhorada	☒	☒	☒
Preparação dos dados	☒	☒	☒
Portabilidade da integração	☒	☒	☒

Conforme se evidencia na Tabela, todas as soluções examinadas satisfazem os critérios de avaliação da Gartner, tornando qualquer uma delas passível de eleição para o desenvolvimento do projeto, mesmo considerando as fragilidades apresentadas. Importa salientar que a escolha da ferramenta de ETL deverá viabilizar a integração com a ferramenta de análise de dados, tema que será abordado na Secção subsequente.

2.5 Ferramentas de Análise de Dados

As ferramentas de análise de dados são plataformas pelas quais qualquer colaborador, mesmo desprovido de conhecimento técnico, consiga modelar, analisar, explorar, gerir e partilhar conhecimento previamente gerado por ferramentas de ETL [42]. Complementarmente, estas ferramentas permitem que as instituições avaliem a conformidade dos seus indicadores chave de desempenho com as ações estratégicas previamente definidas [18].

Conforme indicado pela Gartner, no ano de 2023, a avaliação das ferramentas de análise de dados é realizada com base em doze critérios específicos, os quais desempenharão um papel crucial na seleção da ferramenta de análise de dados a ser empregue na concretização do projeto. Esses critérios são os seguintes [42]:

- *Insights* automatizados – Permite aplicar técnicas de *machine learning* para obtenção de insights valiosos para o utilizador final;
- Catálogo analítico – Disponibiliza conteúdo analítico de fácil acesso para consumo. Neste aspeto é também avaliada a capacidade desse conteúdo ser pesquisável e se a solução oferece recomendações;
- Preparação dos dados – Providencia funcionalidades como a facilidade de arrastar e largar, a integração de dados provenientes de fontes heterógenas, orientadas para o utilizador e, elaboração de modelos analíticos;
- Conectividade com fontes de dados – Integra fontes de dados heterógenas, quer estejam alojadas em plataformas na nuvem ou localmente (*on-premises*);
- Narrativa de dados – Conceder informação crucial para os decisores, de forma inegável, por meio da disposição de *dashboards* conforme uma ordem específica;
- Visualização de dados – Disponibilizar dados sobre ferramentas interativas;
- *Governance* – Capacidade da governação ter controlo sobre o uso e gestão da partilha de informação;
- Consulta em linguagem natural – Possibilitar aos utilizadores a realização das suas pesquisas mediante a utilização, por exemplo, de uma caixa de pesquisa;
- Relatórios – Fornecer relatórios estruturados, passíveis de serem programados e disseminados a um conjunto de utilizadores;
- Integração com ciência dos dados – Permitir utilizar técnicas de *machine learning*, mineração de dados, entre outras;
- Armazenamento de métricas – Disponibilizar aos utilizadores a criação e armazenamento de métricas, permitindo que essas métricas sejam disponibilizadas para posteriores análises;
- Colaboração – Conceder a possibilidade de um vasto conjunto de utilizadores trabalhe simultaneamente na ferramenta.

Segundo os doze critérios mencionados, as soluções foram então divididas nos quatro

quadrantes, definindo a sua posição no mercado atual. Através da Figura 10 é possível observar-se essa distribuição gráfica.

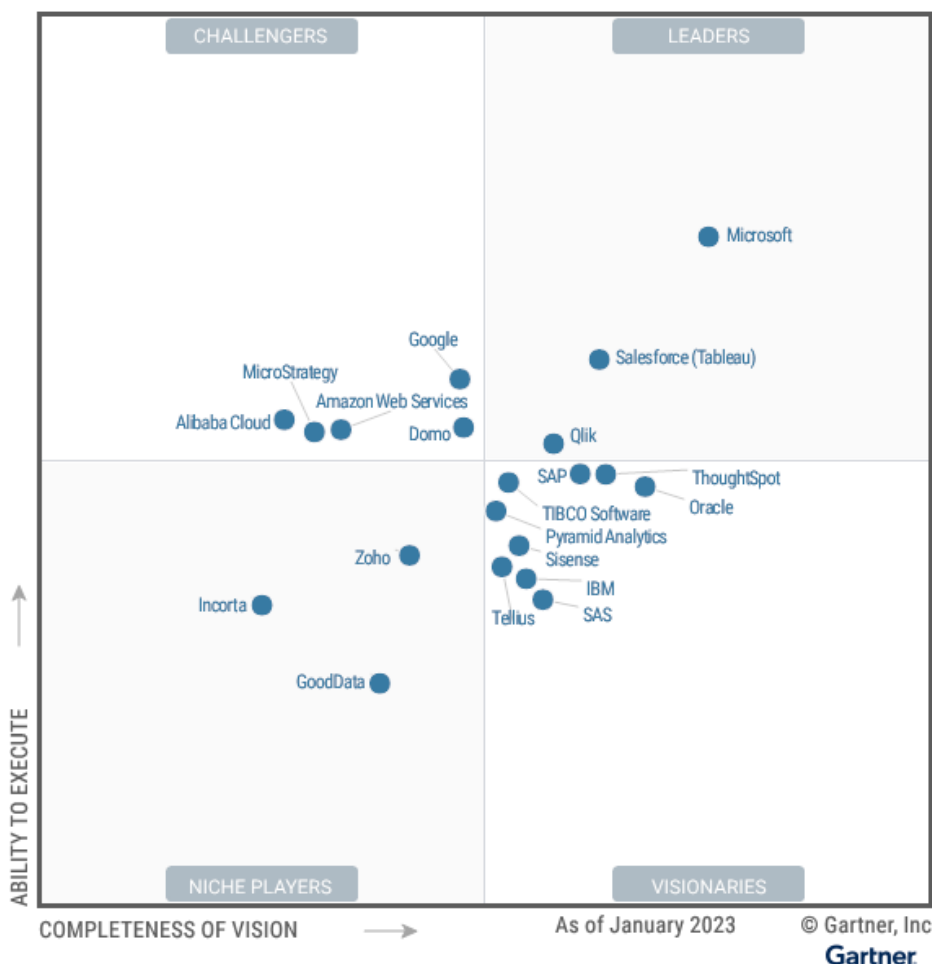


Figura 10 – Quadrante Mágico para Plataformas de Análise e *Business Intelligence* [42]

As ferramentas de análise de dados selecionadas para investigação na revisão literária foram o PowerBI, Qlik Sense e Tableau. A escolha baseou-se no conhecimento prévio do autor, alinhando-se com a presença destas três ferramentas como líderes no quadrante mágico da Gartner para plataformas de análise e BI em 2023 [42].

2.5.1 PowerBI

O Microsoft PowerBI destaca-se como uma ferramenta de análise de dados versátil e personalizável, concebida de forma a facilitar a análise, combinação e pesquisa de dados por parte de qualquer utilizador, recorrendo a *dashboards* interativos e linguagem natural. Adicionalmente, permite a execução do processo de ETL, a criação de métricas, a aplicação de

técnicas de *machine learning* para o consumo dos dados e partilha de painéis desenvolvidos. Este produto, que se encontra disponível em versões gratuita e paga, está acessível tanto na versão desktop como na web, além de possuir uma aplicação móvel [11], [18], [21].

O PowerBI é uma solução que resulta da evolução das funcionalidades como PowerPivot, PowerQuery e PowerView disponibilizadas pelo Microsoft Excel [29]. Comparativamente com ferramentas da mesma génese como o Tableau, esta distingue-se pela fácil integração com outros produtos Microsoft e pela capacidade de se conectar com uma ampla variedade de fontes de dados [18]. Através da Figura 11 é possível consultar um exemplo de *dashboard* desenvolvido em PowerBI.



Figura 11 – Exemplo de *dashboard* desenvolvido em PowerBI [43]

Na análise realizada pela Gartner em 2023, salientou-se a integração bem-sucedida do PowerBI com plataformas como Microsoft 365, Teams e Azure Synapse, evidenciando também a influência determinante do fator preço/valor praticado e a eficácia resultante da combinação das ferramentas, tais como Power Apps e Power Automate, como sendo as principais forças do PowerBI. Porém, apresenta aspetos a melhorar, nomeadamente na gestão de criação e publicação de conteúdos da ferramenta, na integração com outras ferramentas da mesma génese e na exclusividade de implementação no ambiente Azure [42].

2.5.2 Qlik Sense

O Qlik Sense constitui um recurso analítico cujo propósito é gerar relatórios personalizados e *dashboards* intuitivos, dedicados ao contexto empresarial de uma organização. A excelência desta solução evidencia-se através da garantia de segurança ao nível empresarial, da pesquisa inteligente, das visualizações intuitivas de arrastar e largar, da eficaz gestão de conexões de dados, da biblioteca de objetos partilhados, do ambiente ágil de desenvolvimento, do acesso em tempo real, das robustas APIs abertas e normalizadas, assim como pela sua notável história de dados e escalabilidade [44]. A Figura 12 apresenta uma visualização através do Qlik Sense.



Figura 12 – Exemplo de *dashboard* desenvolvido em Qlik Sense [45]

A possibilidade de análises compostas, as capacidades abrangentes de análise de dados, e o facto de ser um produto agnóstico à nuvem, foram identificadas como as mais valias dos produtos da Qlik. Ainda assim, é apontada a ausência de dados circundantes ou de ecossistema de aplicações, ou seja, a desvantagem face a grandes empresas fornecedoras de soluções em nuvem como sendo a principal vulnerabilidade. Para além disso, a prioridade dada à integração de dados e a complexidade de licenciamento são apontadas como outros elementos que contribuem para debilidades das soluções Qlik [42].

2.5.3 Tableau

Tableau, uma ferramenta restrita da Salesforce, define-se como uma solução de análise e visualização de dados acessível e intuitiva. As suas características distintivas, como a inteligência integrada e a otimização de desempenho através da gestão de memória, contribuem para a sua ampla popularidade. Além disso, a capacidade de realizar mapeamentos e apresentar dados geográficos é uma funcionalidade adicional que enriquece a utilidade desta ferramenta [21].

Em comparação com o seu concorrente principal, o PowerBI, o Tableau apresenta exclusivamente uma opção paga que está acessível para avaliação durante um período de catorze dias. Esta oferta concede aos utilizadores acesso a funcionalidades suplementares e beneficia ainda de um programa de suporte técnico sem custos adicionais. Complementarmente, a plataforma disponibiliza uma versão para desktop e nuvem, além de uma aplicação móvel gratuita para partilha de dados [11]. A Figura 13 apresenta um *dashboard* desenvolvido em Tableau.

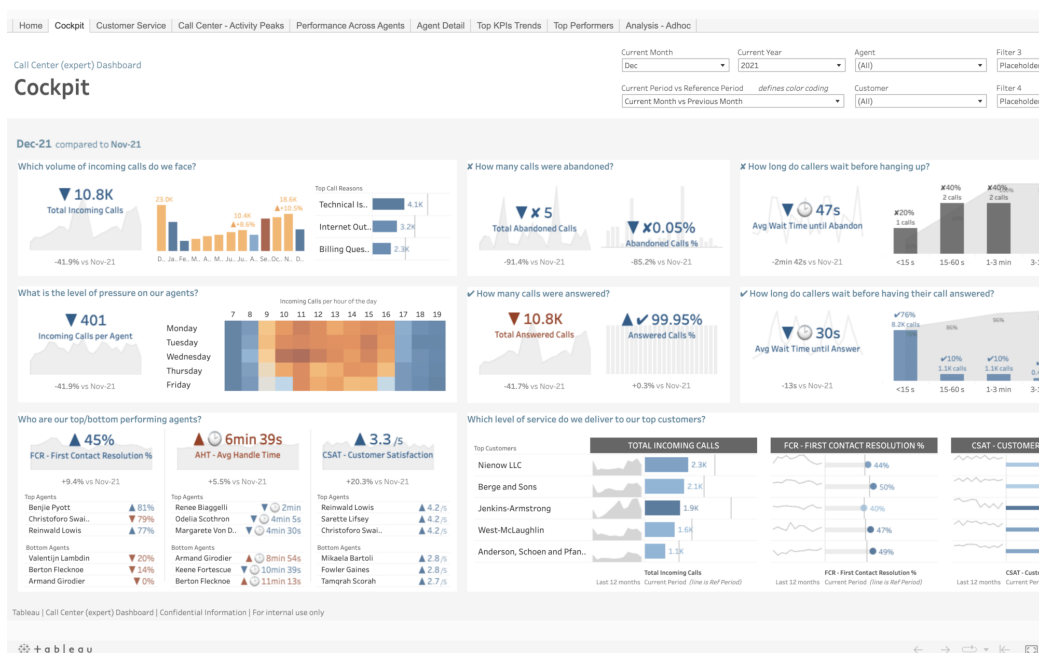


Figura 13 – Exemplo de *dashboard* desenvolvido em Tableau [46]

Conforme indicado pela avaliação da Gartner no ano de 2023, o Tableau destaca-se pelas vantagens que incluem o sólido senso de comunidade à volta da ferramenta, favorecendo a sua evolução, a sua abordagem agnóstica e flexível, assim como a experiência analítica proporcionada. No entanto, algumas fragilidades foram identificadas, como a redução da equipa de suporte face ao crescimento no mercado, a concentração da inovação em Salesforce

e as dificuldades decorrentes da transição para a nuvem, tornando-se desafiante o apoio prestado aos clientes que optam por manter as soluções *on-premises* [42].

2.5.4 Análise Comparativa entre Ferramentas de Análise de Dados

A Tabela 2 expõe uma análise comparativa das ferramentas elegidas para a análise de dados selecionadas, considerando os critérios usados pela Gartner na avaliação deste segmento de ferramentas.

Tabela 2 – Comparação entre ferramentas de análise de dados

Critérios	PowerBI	Qlik Sense	Tableau
<i>Insights</i> automatizados	☒	☒	☒
Catálogo analítico	☒	☒	☒
Preparação dos dados	☒	☒	☒
Conectividade com fontes de dados	☒	☒	☒
Narrativa de dados	☒	☒	☒
Visualização de dados	☒	☒	☒
<i>Governance</i>	☒	☒	☒
Consulta em linguagem natural	☒	☒	☒
Relatórios	☒	☒	☒
Integração com ciência dos dados	☒	☒	☒
Armazenamento de métricas	☒	☒	☒
Colaboração	☒	☒	☒

Após a análise das várias soluções selecionadas, em conjunto com o que é apresentado na Tabela 2, conclui-se que todas as ferramentas de análise de dados possuem características muito semelhantes e atendem aos critérios de avaliação estabelecidos pela Gartner, apesar de algumas fragilidades evidenciadas. Dado que, de acordo com esses critérios, nenhuma delas se destaca face às restantes e todas são passíveis de integração com as ferramentas de ETL

previamente apresentadas, pode-se afirmar que qualquer uma delas é viável para ser escolhida no desenvolvimento da solução.

2.6 Sports Analytics

Elia Morgulev, Ofer H. Azar e Ronnie Lidor caracterizam *Sports Analytics* como um domínio da análise de dados que se dedica à “investigação e modelação do desempenho desportivo, implementando técnicas científicas”[29]. Este campo envolve a integração da gestão de dados históricos com a aplicação de modelos analíticos e utilização de sistemas de informação que auxiliam os decisores na tomada de decisões, visando obter uma vantagem competitiva durante uma partida [13], [29]. A análise desdobra-se em três níveis, a preditiva, prescritiva e a descritiva. A análise preditiva visa antecipar eventos futuros, utilizando técnicas de *machine learning*. A análise prescritiva, por sua vez, tem como objetivo examinar as tendências atuais e previsões, sendo geralmente aplicada em investigação operacional e ciências da gestão para otimizar um sistema. Por último, a descritiva ou *reporting* refere-se à representação do conhecimento por meio de *dashboards*, indicando o que acontece dentro de uma organização, compreendendo tendências e suas causas subjacentes. Este último é o foco central da presente dissertação. A Figura 14 proporciona uma visão gráfica dessas três categorias [47], [48].

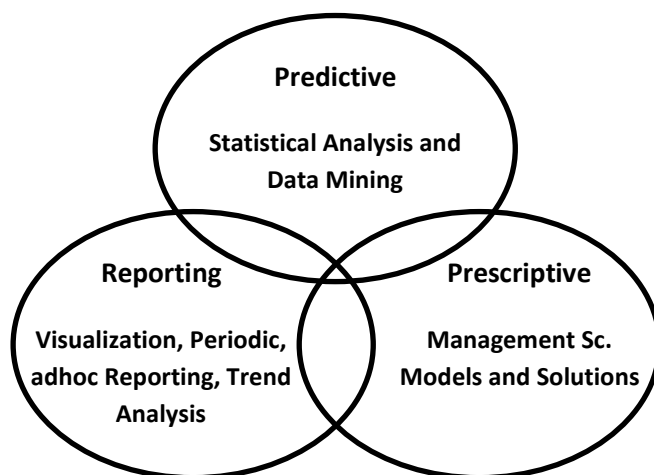


Figura 14 – Categorias de análise de *Sports Analytics* (Adaptado de [47])

A área da *Sports Analytics* tem vindo a crescer rapidamente, apresentando uma ampla gama de métodos que oferecem diversas vantagens, independentemente da perspetiva pela qual são aplicados [29]. Suas funções abrangem a comparação de desempenho entre atletas e respetivas equipas, previsão e correlação do desempenho com base em diversos atributos quantitativos

ou qualitativos, geralmente obtidos a partir de dados biográficos, relatórios de *scouting*, desempenho e, ou médicos [13], [15]. No contexto de um clube desportivo, os treinadores e a equipa técnica, possuindo conhecimento aprofundado de um jogador, agregam valor na constituição da equipa, mudança de dinâmica de uma partida, prevenção de lesões, em negociações de transferência e em outros fatores, tanto qualitativos como quantitativos [29], [49].

Estes dados são recolhidos, padronizados, e analisados por meio de um conjunto de métricas. Por fim, são tomadas decisões fundamentas nessas mesmas análises, aumentando as probabilidades de sucesso, tanto no âmbito desportivo como na gestão organizacional. Esta estrutura, desenvolvida por Elia Morgulev, Ofer H. Azar e Ronnie Lidor (Figura 15), assemelha-se ao processo de BI, particularmente ao processo de ETL [9], [10], [17].

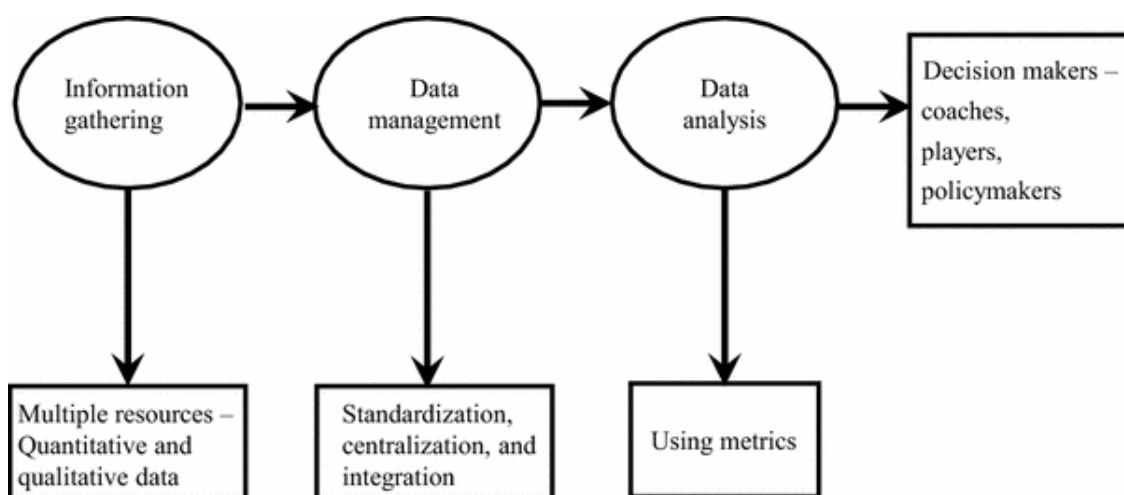


Figura 15 – Estrutura de *Sports Analytics* [9], [10], [17]

A *Sports Analytics* teve a sua primeira incursão no beisebol, onde a utilização de estatísticas avançadas permitiu a seleção criteriosa da constituição inicial de jogadores e a definição de estratégias. Este fenómeno ficou imortalizado no filme "Moneyball", que relata como Billy Beane, diretor geral dos Oakland Athletics, conduziu a equipa a uma série de vitórias notável [11], [15], [50], [51]. Desde então, esta abordagem tem sido amplamente adotada, entre os quais encontram-se os clubes de basquetebol Dallas Mavericks e San Antonio Spurs [52], bem como a seleção germânica de futebol, vencedora do mundial de 2014 [15]. Para além disso, o caso de Kevin Durant, jogador de basquetebol, intensifica a tendência, ao contratar um especialista em análise estatística para melhorar o seu desempenho [21].

Olivier Caya e Adrien Bourdon conceberam uma estrutura que ilustra o contributo que a *Sports Analytics* pode proporcionar às organizações desportivas, mediante a aplicação de ferramentas de *Business Intelligence & Analytics* (BI&A) (Figura 16). Nesta estrutura, são igualmente identificadas todas as partes interessadas, distribuídas em quatro conjuntos distintos [9], [10], [17], [21], [50]:

- Institucional – Categoria principal que engloba as instituições e federações desportivas responsáveis pela gestão do desporto. Nesta estão incluídas a FIFA e a UEFA mencionadas anteriormente;
- Organizacional – Categoria intermédia, que compreende as equipas e os profissionais de desporto, como treinadores, equipa técnica, direção, departamento de análise, entre outros;
- Individual – Categoria composta pelos atletas;
- Externos – Esta categoria abrange um conjunto de indivíduos que não atribuem importância ao valor da utilização de ferramentas de BI&A, mas desempenham um papel fundamental como consumidores principais de dados desportivos. Incluem-se aqui sócios dos clubes, adeptos, empresas de análise de dados, bloggers e sites desportivos, entre outros.

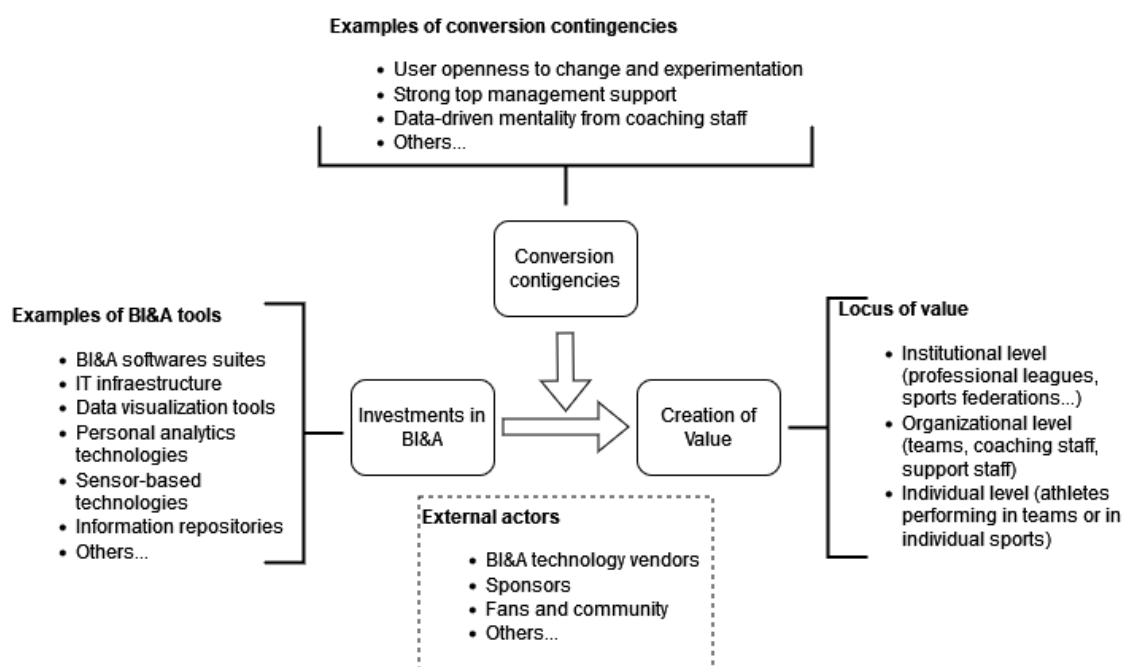


Figura 16 – Proposta de criação de valor no desporto utilizando ferramentas de BI&A por Caya e Bourdon [50]

Nem todas as organizações desportivas possuem recursos humanos e financeiros suficientes para usufruir de ferramentas de *Sports Analytics* [50]. O jogo eletrónico *Football Manager* (FM) e o *Transfermarkt*, um site especializado em avaliações de mercado de jogadores, têm desempenhado um papel significativo, proporcionando ferramentas economicamente acessíveis. Essas plataformas permitem que clubes com recursos financeiros mais limitados tenham a oportunidade de criar valor para suas instituições ao acederem a informações que, de outra forma, só estariam disponíveis através da aquisição de plataformas pagas [6], [53],[6].

2.6.1 Football Analytics

A *Football Analytics*, ramo da *Data Analytics* aplicado ao futebol, subdivide-se em dois meios de análise, a quantitativa e a qualitativa, que por meio da identificação e recolha de indicadores, permitem uma “maior racionalização de padrões e modelos de jogo” [6]. Estes proporcionam aos treinadores, *scouts* e investigadores a obtenção de um maior conhecimento e, conseqüentemente uma melhor prestação de serviços desportivos no domínio do futebol [6], [52]. Além disso, a *Football Analytics* contribui para o desenvolvimento dos atletas, identificando áreas a serem melhoradas e possibilitando a antecipação de comportamentos específicos. Um exemplo disso ocorre na marcação de penalidades, em que o guarda-redes pode estudar previamente para qual lado o marcador provavelmente realizará o remate, com base nas penalidades executadas anteriormente [52].

A análise quantitativa visa abordar questões específicas, como o modo como o adversário marca e sofre golos, bem como a utilização de jogadores em determinadas posições. A análise qualitativa, por sua vez, é realizada através da observação da equipa adversária [10]. Na Tabela 3, expõe-se alguns exemplos de dados recolhidos para análises quantitativas e qualitativas no contexto do futebol.

Tabela 3 – Exemplos de dados recolhidos para análises qualitativas e quantitativas no futebol [10]

Análises Quantitativas		Análises Qualitativas	
Dados Gerais	Dimensões do campo, títulos, treinador e previsão meteorológica	Posicionamento	Posições ocupadas pelos jogadores na primeira parte da construção do adversário, se é mais conservadora ou de alta pressão
Lista do Plantel	Dados biográficos dos jogadores (nome, posição, idade, altura, peso, pé predominante e nacionalidade)	Pontos fortes do adversário	Identificação dos melhores aspetos de forma a preveni-los
Jogadores utilizados (ao minuto)	Todos os minutos jogados por cada jogador, e a posição em que ocuparam, por jogo	Distribuição de bola do guarda-redes	Consiste na opção que o guarda-redes toma para iniciar a construção da equipa (distribuição longa ou curta)
Jogadores mais utilizados	Definição dos jogadores mais utilizados na seleção do onze titular	Lances de bola parada	Forma como a equipa defende e ataca em lances de bola parada (cantos, livres e até lançamentos)
Zonas de golo	Local onde os golos são marcados	Visualizar e registar postura de jogadores e equipa	Observação da postura da equipa e, ou jogador em momentos particulares do jogo (exemplo, sofrer um golo)

No âmbito do *Football Analytics*, destaca-se o exemplo do Liverpool, cujo trajeto foi significativamente influenciado pela nomeação do treinador alemão Jürgen Klopp em outubro de 2015, cuja ligação ao clube terminou no final da presente época desportiva 23/24. Klopp, um técnico que reconhecia a importância das análises no futebol, guiou a equipa à conquista

de diversos troféus, incluindo a UEFA Champions League em 2019 [9], [29]. Este caso, juntamente com outros, será abordado em detalhe na Secção 2.7 dedicada aos casos de estudo [6].

2.6.2 Soluções de Sports Analytics

As soluções oferecidas pela *Sports Analytics* constituem ferramentas “de deteção que fornecem fluxos de dados de alta fidelidade extraídos de cada partida” [53]. Dentro desse contexto, identificam-se dois tipos, os serviços de anotação e os dispositivos de anotação e recolha de dados. Ambos possuem dados relevantes para o desporto analisado, no entanto, os dispositivos de anotação, além de conterem software, incluem também produtos IoT associados, concentrando-se na recolha de índices físicos dos atletas. Por outro lado, os serviços de anotação, centram-se na extração, análise e geração de dados estatísticos desportivos através de vídeo. Importa salientar que o acesso a esses dados requer um pagamento ajustado ao seu valor [53].

De seguida, são apresentados exemplos de dispositivos de anotação, tais como CATAPULT Sports e WHOOP, juntamente com o exemplo de serviço de anotação, Opta Stats Perform. As escolhas destas soluções constituem, segundo a Brand Essence Market Research, três das cinco melhores soluções de *Sports Analytics* no ano de 2023 [54].

2.6.2.1 CATAPULT Sports

A CATAPULT Sports, com sede em Melbourne, Austrália, oferece tecnologia de análise de desempenho desportivo a equipas de várias modalidades, incluindo futebol, beisebol, futebol americano, hóquei no gelo, críquete e basquetebol. A sua oferta compreende três serviços, a análise de vídeo, recrutamento e seleção e monitorização de atletas. No âmbito da análise de vídeo, são disponibilizadas duas soluções, a CATAPULT PRO VIDEO, direcionada para o futebol, e o CATAPULT THUNDER, mais adequado para futebol americano e hóquei. No domínio do recrutamento e seleção, a empresa apresenta duas soluções com a mesma denominação em inglês, SCOUT e RECRUIT. Por último, no campo da monitorização de atletas, destaca-se um conjunto de soluções de IoT, entre as quais se evidenciam a CATAPULT ONE, VECTOR CORE e VECTOR PRO [55].

As soluções de monitorização de atletas são comumente identificadas pelo uso de coletes e/ou braçadeiras, nos quais está integrado um dispositivo de hardware para acompanhar o

desempenho do atleta. Adicionalmente, estas soluções incluem uma aplicação móvel que possibilita o acompanhamento dos mesmos indicadores, independentemente da localização [55].

2.6.2.2 WHOOP

A WHOOP é uma empresa estabelecida em Massachusetts, no estado de Boston, nos Estados Unidos. Esta dedica-se ao desenvolvimento de dispositivos IoT para monitorização da condição física e saúde, direcionados a atletas e entusiastas do fitness. Os produtos da empresa englobam uma pulseira, denominada de WHOOP Band, capaz de registar vários indicadores biométricos, tais como ritmo cardíaco, padrões de sono e recuperação, entre outros [54], [56].

2.6.2.3 OPTA Stats Perform

A Stats Perform, atualmente com sede em Illinois, no estado de Chicago, nos Estados Unidos, adquiriu o nome de OPTA Stats Perform após a aquisição da OPTA [3]. Assim como as soluções já referidas, trata-se de uma empresa de tecnologia e dados desportivos que disponibiliza informações, análises e dados de organizações desportivas, empresas de comunicação social e operadores de apostas [53], [54]. Por meio da aplicação de análises avançadas e *machine learning*, a OPTA Stats Perform processa uma elevada quantidade de dados, fornecendo *insights* valiosos aplicáveis ao recrutamento e desenvolvimento de jogadores, à evolução da estratégia de jogo e à melhoria da experiência dos adeptos [3], [54].

A OPTA Stats Perform destaca-se pela abrangência na cobertura de diversas competições desportivas em todo o mundo, não se limitando apenas ao universo do futebol, conforme referido por [3]. Entre as suas ferramentas de maior relevo, destacam-se [3]:

- ProVision – Ferramenta que permite comparar e analisar jogadores e equipas exaustivamente através de KPI definidos;
- ProPortal – Ferramenta que fornece dados em tempo real às equipas técnicas, capacitando-as para identificar padrões tanto na sua própria equipa quanto no adversário. Sendo de importância crucial para a tomada de decisões durante uma partida;
- ProVideo – Ferramenta com o propósito de identificar o estilo de jogo de uma equipa, analisar taticamente a mesma e avaliar a influência de cada jogador na dinâmica global da equipa.

2.6.2.4 Análise Comparativa entre Ferramentas Sports Analytics

A análise comparativa das soluções de *Sports Analytics*, apresentada na Tabela 4, enfatiza cinco critérios relevantes para os profissionais de futebol, analistas e investigadores. Estes critérios abrangem a capacidade de conduzir análises no contexto do futebol, a adequação ao domínio específico, os custos associados, a presença de aplicativo móvel e a disponibilidade de API.

Tabela 4 – Comparação entre soluções de *Sports Analytics*

Crítérios	CATAPULT Sports	WHOOOP	OPTA Stats Perform
Possibilidade de realizar análises no futebol	Sim	Sim	Sim
Indicado para análises de futebol	Sim	Não	Sim
Custo (Nenhum/Baixo/Médio/Elevado)	Alto	Alto	Médio
Disponibiliza app mobile	Sim	Sim	Não
Disponibiliza API	Sim	Sim	Sim

Após analisar diversas soluções de *Sports Analytics* apresentadas na Tabela 4, é possível concluir que todas elas têm a capacidade de fornecer dados para análises de futebol, assim como disponibilizar dados através de API. Entre as soluções mencionadas, os dispositivos de anotação destacam-se pelo custo mais elevado, resultante da aquisição e/ou aluguer de equipamentos IoT. Por fim, constata-se que a WHOOP não é uma solução adequada para análises desportivas de futebol, e a OPTA Stats Perform não oferece aplicação móvel.

2.6.3 Websites de Estatísticas e Análises

Existem vários sites especializados em fornecer informações sobre eventos desportivos passados e futuros [57]. No entanto, a maioria dessas plataformas consiste em extensas tabelas repletas de dados, tornando difícil encontrar todas as informações relevantes de forma organizada num único local. Isso resulta em ferramentas de baixa usabilidade, o que complica o processo de análise [2].

A introdução de dados é realizada por duas vias, manualmente e, ou através de API [57]. A introdução manual apresenta vulnerabilidades, como possíveis atrasos no acesso aos dados em tempo real e uma alta probabilidade de introdução de erros. Por outro lado, a entrada via API, mantém a informação coerente nessas soluções. No entanto, isso também implica a necessidade de realizar subscrições sobre os dados [2], [57].

A seguir, serão expostos os sites de estatísticas e análises, tais como KickForm, Sofascore, Transfermarkt, que foram classificados por [58] como os três melhores websites para estatísticas e análise desportivas em 2023.

2.6.3.1 KickForm.com

A plataforma KickForm, desenvolvida na Universidade de Munster, Alemanha [59] , constitui uma valiosa fonte de estatísticas, probabilidades e previsões cientificamente fundamentadas, disponibilizando gratuitamente esses dados a apostadores. As estatísticas abrangem as principais ligas de futebol da Alemanha, França, Itália, Inglaterra e Espanha, bem como as segunda e terceira ligas alemãs [58], [59]. Essas informações são atualizadas semanalmente, fornecendo uma variedade de dados relevantes para prever o desfecho de um jogo, incluindo detalhes sobre o confronto direto, a forma atual das equipas, a composição mais recente, entre outros elementos [58]. Além disso, a plataforma oferece uma explicação detalhada do algoritmo utilizado para gerar as previsões [59].

2.6.3.2 Sofascore.com

O Sofascore, uma plataforma líder no conjunto de websites dedicados a estatísticas e análises desportivas, destaca-se como a ferramenta com a mais extensa base de dados estatística, abrangendo mais de 11.000 torneios e mais de 600 ligas em 20 modalidades desportivas, com suporte para 30 idiomas. Originário de Zagreb, Croácia, este site possibilita o acompanhamento em tempo real de eventos desportivos, apresentando diversas estatísticas das equipas ao longo das partidas. Além disso, oferece a visualização de vários indicadores individuais dos jogadores, como golos, dribles, faltas e passes [58].

O Sofascore sobressai ainda pela funcionalidade do mapa de calor, e um centro de jogo constantemente atualizado que mantém os espetadores informados sobre mudanças de posse de bola, remates e cantos. Para complementar, dispõe de um sistema de notificações personalizáveis para cada utilizador [58].

2.6.3.3 Transfermarkt.co.uk

Como previamente abordado, o Transfermarkt emerge como uma plataforma de análise futebolística de extrema relevância nos dias de hoje, desempenhando um papel crucial na obtenção de informações acerca das estimativas de valor de mercado de vários jogadores, sobretudo os mais jovens. Adicionalmente, esta plataforma alemã, sediada em Hamburgo, disponibiliza uma ampla gama de dados históricos sobre as partidas, desempenho de equipas e atletas, os quais exercem uma influência direta no valor apresentado. Além disso, o Transfermarkt conta com módulos dedicados a notícias atuais e rumores de transferências [58], [60].

2.6.3.4 Análise Comparativa entre Websites de Estatísticas e Análises

A Tabela 5 evidencia a comparação analítica entre as diversas soluções de páginas web dedicadas a estatísticas e análises, empregando os mesmos cinco critérios previamente aplicados na avaliação de soluções de *Sports Analytics*.

Tabela 5 – Comparação entre websites de estatísticas e análises no futebol

Critérios	KickForm	Sofascore	Transfermarkt
Possibilidade de realizar análises no futebol	Sim	Sim	Sim
Indicado para análises de futebol	Sim	Sim	Sim
Custo (Nenhum/Baixo/Médio/Elevado)	Nenhum	Nenhum	Nenhum
Disponibiliza app mobile	Sim	Sim	Sim
Disponibiliza API	Não	Não	Não

Com base na Tabela apresentada, é possível inferir que todos os websites analisados são semelhantes em todos os critérios.

2.7 Casos de Estudo

Na presente Secção, serão expostos exemplos concretos de êxito resultantes da aplicação de ferramentas de *Sports Analytics* no contexto do futebol.

2.7.1 Dashboard para “La Liga”

No contexto da La Liga, a principal liga espanhola, o departamento de BI e análise de dados tem vindo a disponibilizar informações de maneira a permitir que os clubes compreendam a situação de mercado e, com base nisso, tomem decisões estratégicas [61].

Com esse propósito, o departamento desenvolveu um conjunto de *dashboards* utilizando o PowerBI, nos quais são apresentados indicadores específicos para o denominado *Sports Business Analytics*. Esses indicadores abrangem diversos aspetos financeiros, incluindo a venda de bilhetes, receitas provenientes de publicidade, custos associados às transações de jogadores, bem como compras e vendas de atletas. Além disso, é possível analisar indicadores relativos à assistência nos estádios, visualização em canais de televisão nacional e internacional, e número de seguidores nas redes sociais. Esses valores podem ser filtrados por clube pertencente à primeira liga espanhola [61].

2.7.2 Football Analytics no Liverpool

O Liverpool FC, conhecido como Reds, foi fundado em 1892 e experimentou seu auge nas décadas de 70 e 80, acumulando 11 títulos entre 1973 e 1990 [57]. Em 2010, o clube enfrentou sérias dificuldades financeiras, levando a Fenway Sports Group, que já possuía os direitos do clube de beisebol Boston Red Sox e havia tentado contratar Billy Bean como diretor-geral em 2002, a adquirir o Liverpool. A partir desse momento, ficou evidente que todas as decisões seriam orientadas por análise de dados [5], [57].

A primeira medida tomada foi a nomeação de Damien Comolli para o cargo de diretor de estratégia, que resultou nas contratações sonantes de Andy Carroll e Luís Suárez por um montante total de noventa e três milhões, além da transferência de Fernando Torres para um dos seus rivais, o Chelsea FC. Adicionalmente, foram adicionadas as aquisições de Stewart Downing, notável pela sua habilidade de drible e criação de oportunidades, e Jordan Henderson, reconhecido pela sua contribuição nos passes, representando 13,6% da sua equipa anterior, o Sunderland. Os restantes três jogadores afirmaram-se como figuras emblemáticas para o clube ao longo das gerações subsequentes, mesmo perante a carência de títulos pelo clube [5], [57].

Em 2015, conforme previamente exposto, o técnico Jurgen Klopp, que já havia demonstrado sua convicção na análise de dados em seus trabalhos anteriores nos clubes Mainz e Dortmund,

foi contratado. A venda de Philippe Coutinho ao Barcelona, por cento e trinta milhões, possibilitou a contratação de Mohamed Salah, um extremo egípcio considerado excedentário pelo Chelsea FC, juntamente com Allison, um guarda-redes brasileiro, e Van Dijk, um defesa central holandês. Essas contratações, apesar de pouco conhecidas, eram adequadas ao estilo de jogo de Klopp, e culminaram na conquista da Liga dos Campeões da UEFA em 2019 e na Premier League em 2020 [5].

2.7.3 Casos de Sucesso de Saint-Gilloise e Brighton Albion

O Brighton e o Saint-Gilloise partilham a característica distintiva de terem ambos como proprietário Tony Bloom, empreendedor e fundador da Starlizard. Esta empresa, especializada em apostas desportivas com base em análise de dados, aplicou esse conhecimento no contexto do futebol [3].

2.7.3.1 Brighton Albion

O Brighton & Hove Albion FC, fundado em 1901, vulgarmente conhecido como Seagulls, um clube inglês com um histórico limitado de conquistas, passou por uma transformação significativa em 2009, quando, na terceira divisão inglesa, foi adquirido por Bloom. Sob a influência do pensamento analítico, ele implementou uma estratégia que visava principalmente identificar e recrutar jovens talentosos que pudessem agregar valor à equipa. A ênfase recaía em jogadores menos conhecidos, muitas vezes provenientes de mercados menos explorados e, consequentemente, mais acessíveis financeiramente [3].

Atualmente, algumas das contratações mais rentáveis incluem Marc Cucurela, Pervis Estupinan, Yves Bissouma e Leandro Trossard. Cucurela, um promissor lateral espanhol adquirido ao Getafe por dezoito milhões durante a temporada 20/21, foi posteriormente transferido na temporada seguinte por sessenta e cinco milhões para o Chelsea FC. Já Bissouma, comprado por dezasseis milhões ao Lille na temporada 18/19, foi vendido na temporada 22/23, para o Tottenham por aproximadamente vinte e nove milhões. Por último, Trossard, formado na academia do Genk na temporada 19/20 foi adquirido por catorze milhões, o qual foi transferido na temporada 22/23 para o Arsenal por vinte e quatro milhões [3].

Na época 22/23, a equipa exibiu um desempenho notável ao assegurar a sexta posição, o que lhe conferiu a oportunidade de estrear-se na fase de grupos da Liga Europa da UEFA. No decorrer da temporada 23/24, a equipa apresenta uma quebra de rendimento, ocupando a

décima primeira posição na primeira liga inglesa, não conseguindo atingir os lugares europeus para a época 24/25 [62]. No contexto da Liga Europa, a equipa destacou-se ao conquistar quatro vitórias em seis jogos possíveis na fase de grupos, sofrendo apenas uma derrota e um empate. Esta notável performance permitiu-lhe garantir o primeiro lugar no grupo, que contava com clubes como Ajax, Marselha e AEK de Atenas, e assim avançar para a próxima fase. A sua participação terminou nos dezasseis avos de final com a derrota por 5-1 contra a Roma no conjunto das duas mãos [3], [63].

2.7.3.2 Union Saint-Gilloise

O Royale Union Saint-Gilloise, um clube belga fundado em 1897, encontrava-se, na temporada 18/19, competindo na segunda liga belga. Foi adquirido por Bloom, que aplicou a metodologia previamente implementada no Brighton Albion. Isso resultou na ascensão do clube na temporada 20/21, que agora disputa o título de campeão belga e procura constantemente o acesso às competições europeias. Por meio dessa abordagem, foi possível identificar e adquirir os direitos dos jogadores, incluindo Casper Nielsen, Dante Vanzeir e Deniz Undav. Este último, proveniente de um clube que militava na terceira divisão alemã, foi vendido ao Brighton por sete milhões de euros na temporada 21/22, após conquistar o título de melhor marcador [3].

Na época 23/24, a equipa terminou a fase regular da liga belga com uma vantagem de sete pontos sobre o segundo classificado, o Anderlecht. No entanto, durante o play-off de campeão belga terminou em segunda posição a um ponto do campeão Club de Brugge, conferindo o acesso às eliminatórias de acesso à Liga dos Campeões para a época 24/25 [64]. Quanto à sua prestação na Liga Europa, o Saint-Gilloise experimentou um destino distinto em comparação com o Brighton. Integrando um grupo teoricamente mais acessível, composto por LASK Linz, Toulouse e Liverpool, a equipa ficou em terceiro lugar, acumulando oito pontos e assegurando apenas a qualificação para a Liga Conferência da UEFA. Contudo, destacou-se ao conquistar uma vitória notável sobre o prestigiado Liverpool FC na última jornada, encerrando assim a fase de grupos de maneira profícua [65]. A sua prestação nas competições europeias terminou nos oitavos de final, com uma derrota por 4-0 diante do gigante turco Fenerbahçe, no conjunto das duas mãos [66].

2.8 Conclusões

Após a análise bibliográfica, tornou-se evidente a existência de diversas arquiteturas e tecnologias aplicáveis à realização de um projeto de BI. Constatou-se, igualmente, que esse processo engloba três fases distintas: o ETL, a criação de um Data Mart e o desenvolvimento de relatórios e dashboards. Estas etapas convertem dados em conhecimento e, subsequentemente, em sabedoria, sobre qual é realizado processo decisório. Além disso, foram examinadas diversas ferramentas para cada uma destas fases, e apesar de cada uma apresentar os seus pontos positivos e negativos, todas se revelaram aptas para a implementação do mencionado projeto.

Foi constatado que a *Sports Analytics* representa uma área específica de análise de dados dedicada à investigação dos dados no contexto desportivo, subdividindo-se em três abordagens: preditiva, descritiva e prescritiva. Para além disso, revela-se crucial destacar a proposta apresentada sobre o valor das ferramentas de BI&A nesse mesmo domínio, proporcionando benefícios nos âmbitos institucional, organizacional, individual e externo, como comprovado pelos casos de estudo abordados. No mesmo contexto, foram exploradas diversas soluções de *Sports Analytics*, agrupadas em dois conjuntos: serviços de anotação e dispositivos de anotação, e websites de estatísticas e análises, apresentando notáveis diferenças em termos de abrangência, disposição de informação e propósito.

3 Análise e Design

No presente capítulo, apresenta-se o website de onde se originam os dados utilizados na presente dissertação, juntamente com o seu modelo relacional e o modelo de domínio associado. São descritos os atores do sistema, assim como os requisitos funcionais e não funcionais da solução a desenvolver. Estabelecem-se os indicadores de desempenho a analisar e apresenta-se o modelo dimensional proposto. Finalmente, detalha-se a arquitetura do sistema, incluindo o processo de mapeamento e transformação de dados.

3.1 FootyStats

A FootyStats é um website poderoso e abrangente para os entusiastas e profissionais de futebol que procuram aceder a dados estatísticos detalhados e atualizados sobre competições, equipas e jogadores. Além disso, esta plataforma faculta uma Interface de Programação de Aplicações (API) possível de ser subscrita, a qual concede acesso a uma vasta panóplia de estatísticas de jogos provenientes de mais de 1500 ligas de futebol em todo o mundo. A API agiliza a integração eficiente dos dados estatísticos diretamente em aplicações, facilitando a análise aprofundada e a visualização dos padrões e tendências no desporto mais popular do mundo [67].

Ao oferecer um vasto leque de *endpoints*, a API da FootyStats permite aos utilizadores aceder a informações como resultados de jogos, estatísticas de equipas como desempenho em casa e fora, média de golos marcados e sofridos, entre outras, dados individuais de jogadores e até previsões estatísticas para jogos futuros. A riqueza de dados disponíveis possibilita uma análise

profunda e personalizada, sendo uma mais-valia para quem procura *insights* valiosos no contexto do futebol.

Adicionalmente, a API encontra-se documentada de forma clara e organizada, facilitando o desenvolvimento e a implementação por parte de programadores e analistas desportivos. Com a FootyStats API, os utilizadores têm a capacidade de explorar e utilizar dados estatísticos de futebol de forma flexível e abrangente, abrindo portas para novas abordagens na análise e compreensão deste desporto apaixonante.

3.2 Modelo Relacional

O conceito de base de dados relacional, concebido por Edgar Frank Codd, um programador matemático da IBM nas décadas de 1960 e 1970, permanece até aos dias de hoje como a pedra angular da engenharia de bases de dados. Este modelo é definido pela presença de uma série de relações (ou tabelas) distintas, cada uma contendo vários atributos claramente identificados, com o propósito de facilitar a interpretação dos valores de um tuplo (ou registo) [68]. Na figura 17 é exibido o modelo relacional de dados acessível através da API da FootyStats.

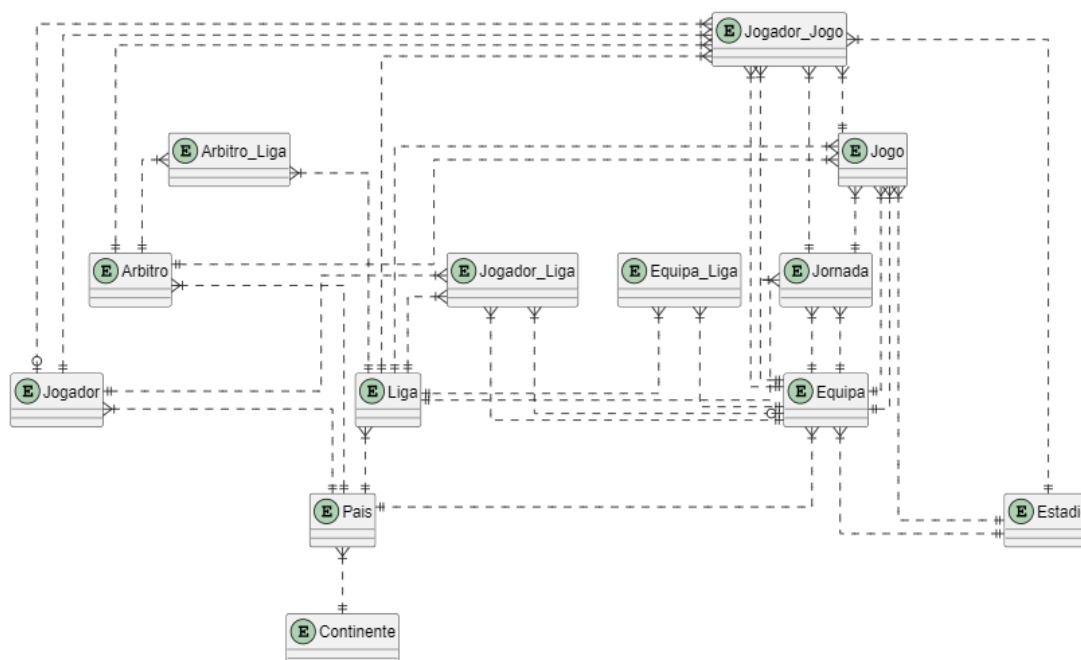


Figura 17 – Modelo de Relacional

No Anexo 2 são apresentadas ao detalhe as diversas tabelas constituintes do modelo relacional, especificando os seus atributos, as descrições correspondentes e os respetivos tipos de dados.

3.3 Modelo de Domínio

Um modelo de domínio constitui uma forma de representação de conhecimento de uma área de aplicação, desempenhando um papel de consenso dentro da comunidade especializada nesse mesmo tema [69]. Neste são representadas as classes conceptuais de negócio, bem como as relações entre elas. Para esta representação foi utilizada a linguagem de modelação unificada (UML), a linguagem indicada para “especificar, visualizar, construir e documentar os artefactos dos sistemas de software” [70].

Como já referido na Secção 1.2 Contexto, o âmbito desta dissertação centra-se no futebol, com ênfase especial no cenário europeu e todos os elementos que o envolvem. Neste sentido, foram examinados dados relacionados com ligas, jogos, equipas e jogadores, visando a extração de conclusões que possam conferir vantagens competitivas. Dada a vasta quantidade de dados disponibilizados pela API da FootyStats, optou-se por limitar o foco da dissertação apenas às dez principais competições de cada país, conforme a classificação atual da UEFA (Tabela 6).

Tabela 6 – Dez países europeus mais bem cotados na UEFA [71]

País	Coeficiente Atual (Pontos)	Nome da 1ª Liga	Campeão Atual	Clube com maior Contributo Atualmente
Inglaterra	104.303	Premier League	Manchester City FC	Manchester City FC
Itália	90.284	Série A	FC Inter de Milão	Associazione Sportiva Roma
Espanha	89.489	La Liga	Real Madrid CF	Real Madrid CF
Alemanha	86.624	Bundesliga	Bayer 04 Leverkusen	FC Bayern Munique
França	66.831	Ligue 1	Paris Saint-Germain FC	Paris Saint-Germain FC
Holanda	61.300	Eredivise	PSV Eindhoven	Amsterdamsche FC Ajax
Portugal	56.316	Primeira Liga	Sporting Clube de Portugal	Sport Lisboa e Benfica
Bélgica	48.800	Jupiler Pro League	Club Brugge Koninklijke Voetbalvereniging	Club Brugge Koninklijke Voetbalvereniging
Turquia	38.600	Spor Toto Super League	Galatasaray Spor Kulübü	Fenerbahçe Spor Kulübü
Escócia	36.050	Scottish Premiership	Celtic FC	Rangers FC

- Evento: Entidade representada pelas ações realizadas por um jogador no decorrer de um jogo;
- Jogador: Entidade que representa os praticantes de futebol;
- Jogo: Entidade que representa o confronto entre duas equipas;
- Jornada: Entidade caracterizada por um agrupamento de jogos numa etapa da época;
- Liga: Entidade que representa uma competição anual;
- País: Entidade que representa a nação a que uma liga, jogador, árbitro e equipa pertencem;
- Treinador: Entidade responsável pela gestão operacional de um clube;
- Treinador-Adjunto: Entidade responsável pelo auxílio na gestão operacional de um clube.

3.4 Engenharia de Requisitos

A engenharia de requisitos é o processo através do qual se interage com os intervenientes do sistema, tendo como o objetivo identificar os atores do sistema e os requisitos funcionais e não funcionais que definem de forma precisa as funcionalidades necessárias da solução. Desta forma, garante-se a satisfação de todas as necessidades das partes interessadas [72].

No âmbito deste processo de engenharia de requisitos, foi realizado um questionário junto de membros da equipa técnica do Rans Nusantara FC [73], conforme apresentado no Anexo 1, com o propósito de proporcionar um enquadramento geral e identificar os requisitos associados à solução a desenvolver.

3.4.1 Atores de Sistemas

No campo da engenharia de requisitos, um ator de sistema assume o papel de uma entidade que interage com o sistema. A identificação de cada um dos atores é fundamental para o êxito do sistema, pois são eles os principais intervenientes e destinatários do mesmo. De seguida é apresentada a Tabela 7 que apresenta esses mesmos atores de sistema.

Tabela 7 – Atores do Sistema

Ator	Descrição
Administrador	Entidade responsável pela administração da solução, na qual se insere a atribuição de acessos.
Treinador	Entidade encarregue de liderar uma equipa, definir estratégias e a formação de jogo, planejar treinos e avaliar os atletas.
Treinador Adjunto	Entidade compreendida como uma extensão do Treinador, com a sua função principal centrada em fornecer uma perspetiva alternativa ao Treinador principal. Em algumas circunstâncias pode também ser encarregue de uma componente específica durante as sessões de treino.
Analista	Entidade encarregue da recolha, processamento, análise e disponibilização de uma variedade de informações relevantes para os Treinadores e seus Assistentes.

3.4.2 Requisitos

Os requisitos delineiam as capacidades que o sistema necessita de incorporar. Estes dividem-se em dois tipos: os requisitos funcionais, que são descritos pelos intervenientes do sistema, e os não funcionais, os quais são geralmente considerados implícitos e cuja não satisfação pode comprometer a adoção adequada da solução em questão. Na definição dos requisitos, adotou-se o modelo de classificação dos atributos de qualidade de software FURPS+, onde cada letra do acrónimo corresponde a um tipo específico de requisito [74]. A avaliação de todos os requisitos será conduzida através de um questionário detalhado, conforme descrito na Secção 5.2 - Avaliação da Solução.

3.4.2.1 Requisitos Funcionais

Os requisitos funcionais, identificados pela letra F no acrónimo FURPS+, são expressos através de um Diagrama de Casos de Uso em UML. Este diagrama ilustra as ações acessíveis, correspondentes às funcionalidades do sistema, as quais podem ser realizadas por múltiplos intervenientes. A Figura 19 apresenta esta representação esquemática.

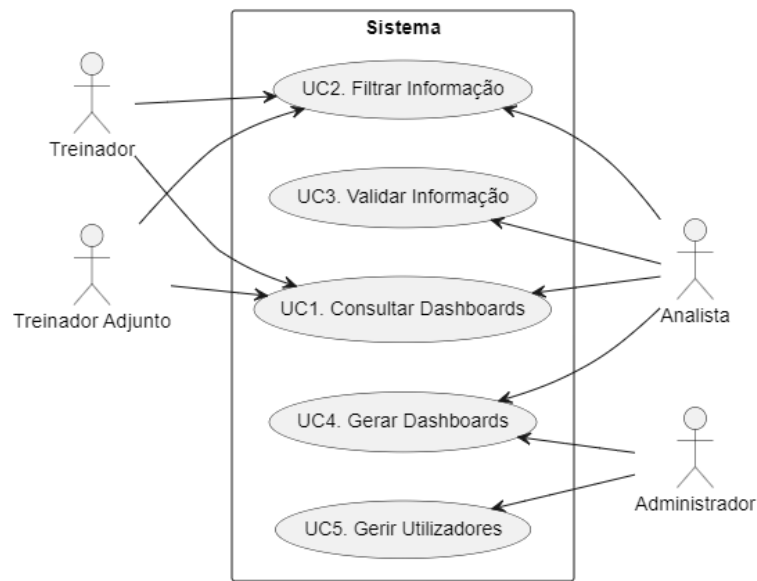


Figura 19 – Diagrama de Casos de Uso

A seguir, são apresentados os diagramas de sequência de sistema de cada caso de uso acima identificado. Estes ilustram todos os passos envolvidos nas interações entre os atores e o sistema.

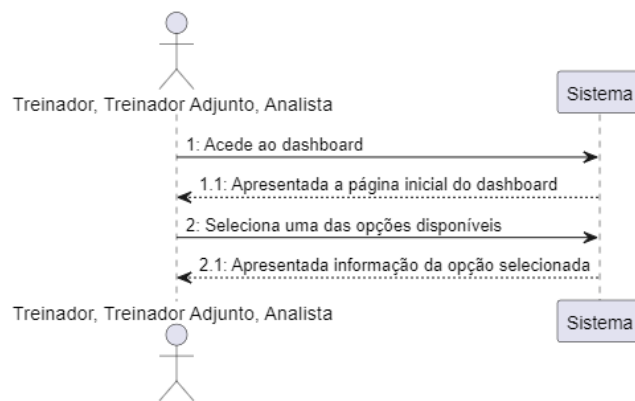


Figura 20 – Diagrama de Sequência de Sistema (Consultar Dashboards)

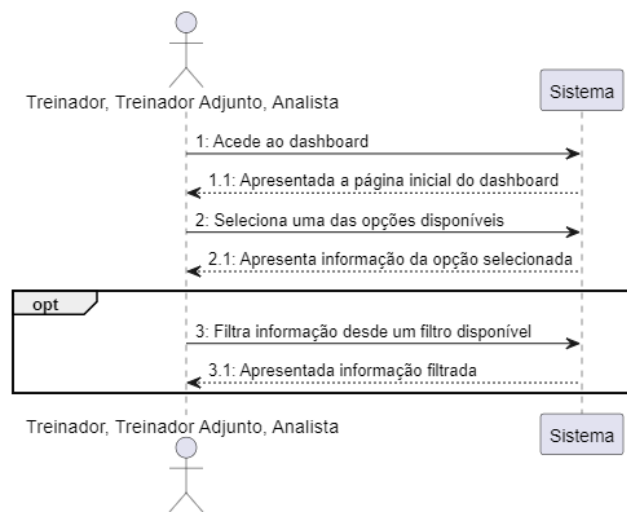


Figura 21 – Diagrama de Sequência de Sistema (Filtrar Informação)

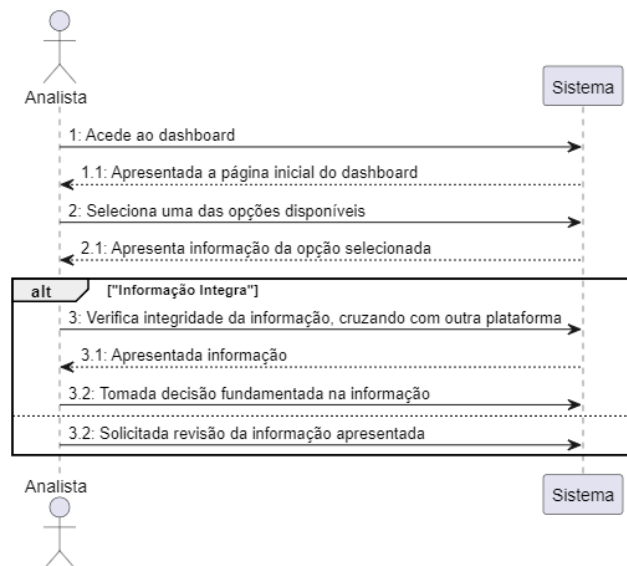


Figura 22 – Diagrama de Sequência de Sistema (Validar Informação)

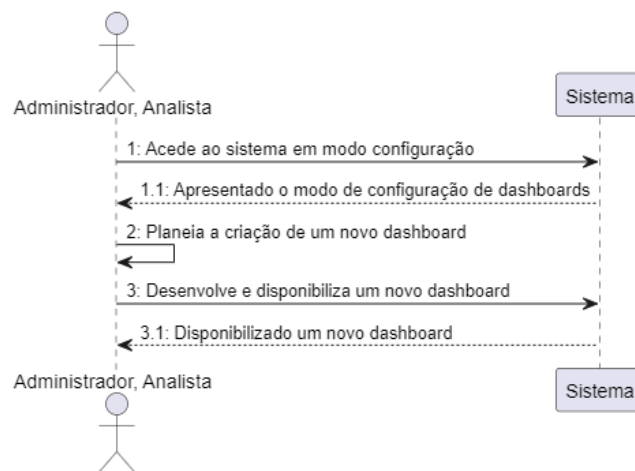


Figura 23 – Diagrama de Sequência de Sistema (Gerar Dashboards)

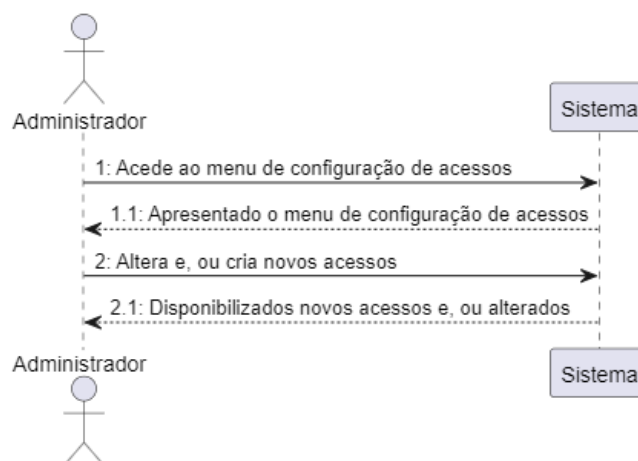


Figura 24 – Diagrama de Sequência de Sistema (Gerir Utilizadores)

3.4.2.2 Requisitos não Funcionais

Os requisitos não funcionais, como anteriormente abordado, descrevem o comportamento esperado da solução, independentemente das funcionalidades específicas. Esses requisitos são delineados pelos atributos do modelo FURPS+, excluindo o componente funcional (letra F). A seguir, são detalhadas as descrições de cada um dos restantes caracteres [75].

- Usabilidade (*Usability*): Atributo encarregado de avaliar tanto a estética quanto a consistência da interface;
- Confiabilidade (*Reliability*): Atributo que avalia a integridade do software, tolerância a falhas, capacidade de recuperação e exatidão dos cálculos efetuados;

- Desempenho (*Performance*): Atributo que avalia o desempenho da solução, entre os quais encontram-se o tempo de resposta, consumo de recursos e disponibilidade da aplicação;
- Suportabilidade (*Supportability*): Atributo que abrange diversos outros requisitos como testabilidade, adaptabilidade, manutenção, compatibilidade, configurabilidade, estabilidade, escalabilidade, etc;
- + (outros): Engloba as múltiplas restrições provenientes de diferentes requisitos, abrangendo aspetos de desenho, implementação, interface e físicos.

Após a exposição das características gerais contempladas pelo modelo em questão, procede-se agora à definição dos requisitos não funcionais pertinentes ao projeto.

Usabilidade – O sistema deve ser intuitivo para que os utilizadores, que não têm conhecimentos técnicos, consigam utilizá-lo com a mesma facilidade que aqueles que os têm. Este critério está alinhado com o princípio mencionado pela letra U (*Usability*).

Integridade dos Dados – Para que as decisões sejam tomadas com maior acuidade e fundamentação, é crucial que os dados se apresentem íntegros e fiáveis. Este critério está alinhado com o princípio da confiabilidade (*Reliability*).

Desempenho – O sistema necessita de processar e exibir os dados com um tempo de resposta inferior a um minuto, uma vez que é essencial que os profissionais de futebol tenham acesso aos mesmos a qualquer momento. Este critério está alinhado com o princípio mencionado pela letra P (*Performance*).

Compatibilidade – O sistema deve ser compatível com variados dispositivos, como telemóveis, tablets, portáteis, entre outros. Este critério está alinhado com o princípio da suportabilidade (*Supportability*).

Modularidade – O sistema deve ser composto por vários módulos independentes entre si, reduzindo o impacto aquando da alteração e, ou remoção de um deles. Este critério está alinhado com o princípio da suportabilidade (*Supportability*).

Histórico – O sistema deve permitir apresentar e preservar informação histórica, uma vez que os dados apresentados no passado poderão servir para apoiar decisões atuais. Este critério está alinhado com a categoria +, no qual se enquadram todos os outros requisitos.

3.5 Definição de Indicadores de Performance

Neste projeto, foi fornecida informação com o intuito de delinear os aspetos de negócio prioritários a considerar na análise do contexto inserido. Como já referido, esta recolha de informação teve lugar sobretudo através de um questionário de levantamento de requisitos dirigido aos profissionais de futebol que realizam análises de dados constantemente, os quais possuem o conhecimento necessário para determinar as métricas relevantes para a solução. Deste levantamento de requisitos, foi claramente evidenciado que a análise de dados realizada para preparar futuros encontros centra-se essencialmente em duas perspetivas, os dados enquanto coletivo (equipa) e os dados individuais (jogador). Assim, a partir do questionário, foram identificados os seguintes indicadores:

- **Acerca do Jogo**
 - Número Total de Golos da Equipa Visitada (homeGoalCount)
 - Número Total de Golos da Equipa Visitante (awayGoalCount)
 - Número Total de Remates à Baliza da Equipa Visitada (team_a_shotsOnTarget)
 - Número Total de Remates à Baliza da Equipa Visitante (team_b_shotsOnTarget)
 - Número Total de Remates Fora do Alvo da Equipa Visitada (team_a_shotsOffTarget)
 - Número Total de Remates Fora do Alvo da Equipa Visitante (team_b_shotsOffTarget)
 - Número Total de Remates da Equipa Visitada (team_a_shots)
 - Número Total de Remates da Equipa Visitante (team_b_shots)
 - Número Total de Faltas da Equipa Visitada (team_a_fouls)
 - Número Total de Faltas da Equipa Visitante (team_b_fouls)
 - Percentagem de Posse de Bola da Equipa Visitada (team_a_possession)
 - Percentagem de Posse de Bola da Equipa Visitante (team_b_possession)
 - Número Total de Cartões da Equipa Visitada (team_a_cards_num)
 - Número Total de Cartões da Equipa Visitante (team_b_cards_num)
 - Número Total de Ataques Perigosos da Equipa Visitada (team_a_dangerous_attacks)
 - Número Total de Ataques Perigosos da Equipa Visitante (team_b_dangerous_attacks)
 - Número Total de Ataques da Equipa Visitada (team_a_attacks)
 - Número Total de Ataques da Equipa Visitante (team_b_attacks)

- Número Total de Golos Marcados por Grande Penalidade da Equipa Visitada (team_a_penalty_goals)
- Número Total de Golos Marcados por Grande Penalidade da Equipa Visitante (team_b_penalty_goals)

- **Acerca do Jogador**
 - Número Total de Minutos Jogados do Jogador na Época (minutes_played_overall)
 - Número Total de Jogos Disputados do Jogador na Época (appearances_overall)
 - Número Total de Golos Marcados do Jogador na Época (goals_overall)
 - Número Total de Jogos sem Sofrer Golos da Equipa Deste Jogador nos Jogos que Participou na Época (clean_sheets_overall)
 - Número Total de Golos Sofridos Enquanto o Jogador Esteve em Campo na Época (conceded_overall)
 - Número Total de Assistências do Jogador na Época (assists_overall)
 - Número Total de Golos Marcados através de Grandes Penalidades do Jogador na Época (penalty_goals)
 - Golos Envolvidos (Golos + Assistências) por 90 Minutos do Jogador na Época (goals_involved_per_90_overall)
 - Golos Marcados por 90 Minutos do Jogador na Época (goals_per_90_overall)
 - Minutos por Golo Marcado do Jogador na Época (min_per_goal_overall)
 - Golos Sofridos por 90 Minutos do Jogador na Época (conceded_per_90_overall)
 - Minutos por Golo Sofrido do Jogador na Época (min_per_conceded_overall)
 - Número Total de Cartões Amarelos do Jogador na Época (yellow_cards_overall)
 - Número Total de Cartões Vermelhos do Jogador na Época (red_cards_overall)
 - Número Médio de Minutos do Jogador por Jogo na Época (min_per_match)
 - Minutos por Assistência do Jogador na Época (min_per_assist_overall)

- **Acerca da Equipa**
 - Número Total de Golos Marcados (seasonScoredNum_overall)
 - Número Total de Golos Sofridos (seasonConcededNum_overall)
 - Diferença Golos (seasonGoalDifference_overall)
 - Número Total de Vitórias (seasonWinsNum_overall)
 - Número Total de Empates (seasonDrawsNum_overall)

- Número Total de Derrotas (seasonLossesNum_overall)
- Número Total de Jogos sem Sofrer Golos (seasonCS_overall)
- Média de Remates à Baliza por Jogo (shotsOnTargetAVG_overall)
- Média de Remates Fora do Alvo por Jogo (shotsOffTargetAVG_overall)
- Média de Posse de Bola por Jogo (possessionAVG_overall)
- Média de Ataques Perigosos (dangerous_attacks_avg_overall)
- Média de Ataques (attacks_avg_overall)
- Número Total de Cartões que a Equipa Recebeu (cards_for_overall)
- Número de Penaltis Marcados (penalties_scored_overall)
- Número de Remates por Golo Marcado (shots_per_goals_scored_overall)
- Taxa de Conversão de Remates (shot_conversion_rate_overall)

3.6 Metodologia

A metodologia de modelação do armazém de dados seguida no âmbito da presente dissertação é a metodologia de Kimball. Esta baseia-se em quatro passos principais, determinar a área de negócio, definir a granularidade, ou seja, o nível de detalhe, apurar dimensões e factos a puderem ser analisados, conforme se pode verificar de seguida. A escolha desta abordagem é mencionada na Secção 3.8.

Área de Negócio: Estatísticas de futebol.

Granularidade:

- Estatísticas por jogador, por liga, por equipa e por ano
- Estatísticas por equipa, por liga e por ano
- Estatísticas por evento, por momento do evento, por jogador e por jogo
- Estatísticas por jogo, por jogador, por liga, por equipas (casa e fora), e por data

Dimensões:

- Ligas (DimLeague)
- Jogadores (DimPlayer)
- Estádios (DimStadium)
- Equipas (DimTeam)

- Eventos (DimEvent_Match)
- Minutos do Evento (DimTime_Match)
- Países (DimCountry)
- Jogos (DimMatch)
- Datas (DimDate)

Factos: Apresentados ao detalhe na secção 3.4 Definição de Indicadores de Performance.

Na Figura 25 é ilustrado o modelo multidimensional.

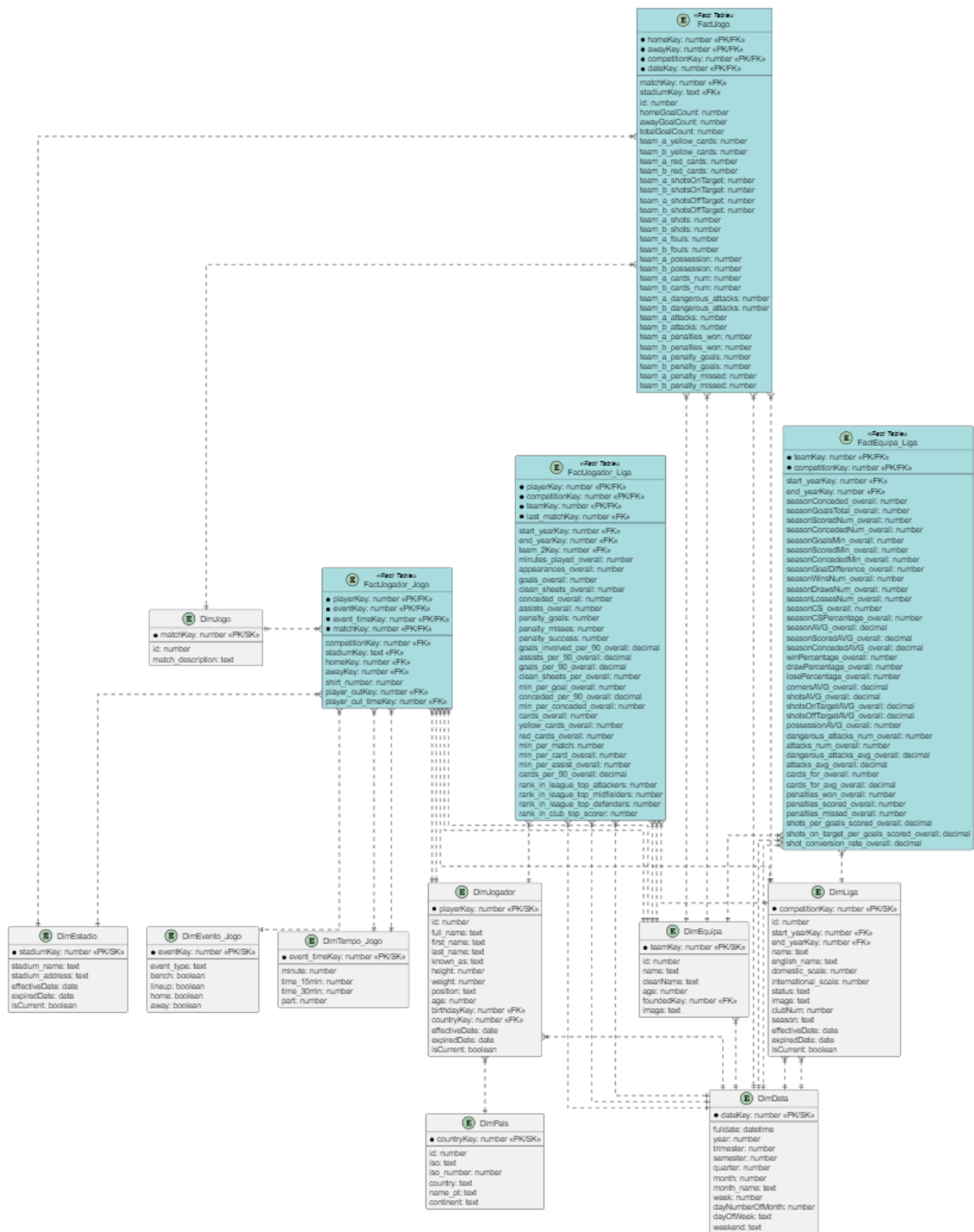


Figura 25 – Modelo Dimensional

3.7 Modelo Multidimensional

O modelo dimensional proposto, juntamente com suas dimensões e tabelas de facto correspondentes, destacadas a azul, pode ser observado na Figura 25. Este diagrama não só

exibe várias tabelas, mas também detalha alguns dos seus atributos e seus tipos, incluindo as chaves primárias. O referido diagrama inclui as seguintes tabelas:

- Tabelas de Dimensão:
 - DimLiga: Dimensão que reúne todos os dados das várias competições ao longo das diferentes temporadas;
 - DimEstadio: Dimensão que reúne todos os dados sobre os estádios nos quais os jogos de futebol são realizados;
 - DimEquipa: Dimensão que reúne todos os dados sobre as equipas que participam nas competições;
 - DimPais: Dimensão que reúne todos os dados sobre os países e respetivos continentes dos atletas;
 - DimJogador: Dimensão que reúne todos os dados sobre os atletas dos diversos clubes;
 - DimData: Dimensão que reúne todos os dados temporais;
 - DimJogo: Dimensão que reúne os dados sobre as partidas;
 - DimTempo_Jogo: Dimensão que contém os minutos dos eventos durante uma partida;
 - DimEvento_Jogo: Dimensão que contém os tipos de eventos possíveis durante um jogo.
- Tabelas de Facto:
 - FactJogador_Jogo: Tabela que contém as principais ações de um jogador durante um jogo, tais como como golos, cartões, entre outros;
 - FactJogador_Liga: Tabela que contém o desempenho de um atleta durante uma temporada;
 - FactEquipa_Liga: Tabela que contém o desempenho de um clube durante uma temporada;
 - FactJogo: Tabela que contém o desempenho das equipas num jogo.

3.8 Arquitetura da Solução

Com base nas arquiteturas propostas na Secção 2.3, foi selecionada a abordagem de Ralph Kimball para a implementação da solução proposta. A escolha desta abordagem deve-se à concentração atual dos dados num único processo de negócio, resultando num armazém de

dados, o que facilitará a gestão da solução por uma equipa reduzida, além da agilidade de implementação proporcionada por essa metodologia, alinhando-se com as diversas vantagens oferecidas por essa abordagem. Assim, em conformidade com o exposto na Subsecção anterior, planeia-se criar um Armazém de Dados (AD) composto por quatro tabelas de fatos: FactJogo e FactJogador_Jogo, FactJogador_Liga e FactEquipa_Liga.

Ao considerar a implementação desta abordagem, é relevante destacar que ela atende a um dos requisitos não funcionais, mencionado na lista apresentada na Subsecção 3.4.2 Requisitos não Funcionais: a modularidade. Isso possibilitará a introdução de novos componentes no futuro, de acordo com as necessidades dos profissionais de futebol ou do clube, sem interferir na solução existente.

A arquitetura consiste numa componente de recolha de dados de uma fonte, uma REST API, que posteriormente são armazenados numa *Staging Area* (SA). Após isso, são submetidos a um processo de transformação para, em seguida, serem carregados no AD. Uma vez concluído esse processo, uma ferramenta de análise de dados será utilizada para construir *dashboards* acessíveis ao conjunto de utilizadores identificados na Subsecção 3.4.1 como atores do sistema. Na Figura 26, é possível visualizar a arquitetura da solução proposta.

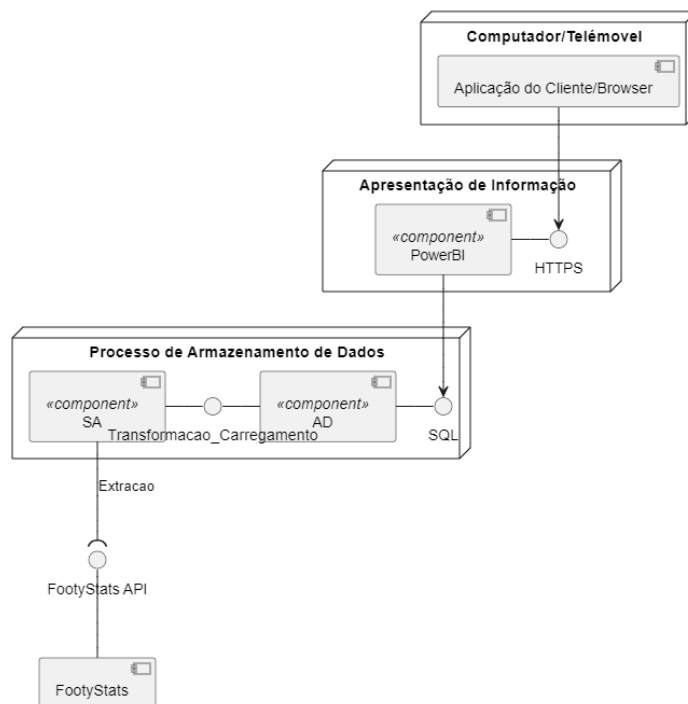


Figura 26 – Arquitetura da Solução

3.9 Mapeamento e Transformação de Dados

A relação entre os atributos da fonte original de dados e o seu destino final é estabelecida por meio de uma tabela de mapeamento lógico de dados. Além do mapeamento direto entre a SA e o AD, são consideradas outras informações cruciais, tais como as transformações planejadas para cada atributo. Estas transformações podem variar desde a alteração do tipo de dados, como de int para nvarchar, até à realização de cálculos para obtenção de informação derivada, assim como a determinação do tipo de SCD (Slow Changing Dimension) aplicável. Esta tabela é essencial não apenas para possibilitar a correspondência e manipulação dos dados brutos, mas principalmente para garantir a consistência e a capacidade de análise dos mesmos. Através da Tabela 8 é possível verificar um excerto dos mapeamentos e transformações de dados no contexto do projeto. No Anexo 3 é apresentada a tabela de mapeamento lógico de dados na sua totalidade.

Tabela 8 – Tabela de Mapeamento Lógico de Dados (Excerto)

Destination Table					Source Table				Transformations
Table Name	Column Name	Data Type	Table Type	SCD Type	Database	Table Name	Column Name	Data Type	
FactJogador_Jogo	playerKey	int	Fact	N/A	Staging Area, Data Mart	DimJogador	playerKey	int	Primary Key where SA_BI_FOOTBALL_1210129. MatchesPlayers.playerID = DM_BI_FOOTBALL_1210129 .DimPlayer.id
FactJogador_Jogo	eventKey	int	Fact	N/A	Staging Area,	DimEvent_Match	eventKey	int	Primary Key where

					Data Mart				SA_BI_FOOTBALL_1210129.MatchesPlayers.event_type = DM_BI_FOOTBALL_1210129.DimEvent_Match.event_type
FactJogador_Jogo	event_timeKey	int	Fact	1	Staging Area	Jogador_Jogo	event_time	nvarchar (20)	<p>Primary Key</p> <pre> SELECT CASE WHEN CHARINDEX('+', event_time) > 0 THEN CAST(SUBSTRING(event_time, 1, CHARINDEX('+', event_time) - 1) AS INT) + CAST(SUBSTRING(event_time, CHARINDEX('+', event_time) + 1, LEN(event_time) - CHARINDEX('+', event_time) - 1) AS INT) ELSE CAST(SUBSTRING(event_time, 1, LEN(event_time) - 1) AS INT) END AS event_time FROM MatchesPlayers </pre>
FactJogador_Jogo	matchKey	int	Fact	N/A	Staging Area,	FactMatches	id	int	Primary Key

					Data Mart				where SA_BI_FOOTBALL_1210129. MatchesPlayers.matchID = SA_BI_FOOTBALL_1210129. FactMatches.id
FactJogador_Jogo	competitionKey	int	Fact	N/A	Staging Area, Data Mart	DimLeague	competitionKey	int	Foreign Key where SA_BI_FOOTBALL_1210129. MatchesPlayers.competitionID = DM_BI_FOOTBALL_1210129. .DimLeague.id
FactJogador_Jogo	stadiumKey	int	Fact	N/A	Staging Area, Data Mart	DimStadium	stadiumKey	int	Foreign Key where SA_BI_FOOTBALL_1210129. MatchesPlayers.stadium_name = DM_BI_FOOTBALL_1210129. .DimLeague.stadium_name
FactJogador_Jogo	awayKey	int	Fact	N/A	Staging Area, Data Mart	DimTeam	teamKey	int	Foreign Key where SA_BI_FOOTBALL_1210129. MatchesPlayers.awayID = DM_BI_FOOTBALL_1210129. .DimTeam.id
FactJogador_Jogo	shirt_number	int	Fact	N/A	Staging Area	Jogador_Jogo	shirt_number	int	

FactJogador_Jogo	player_outKey	int	Fact	N/A	Staging Area, Data Mart	DimJogador	playerKey	int	Foreign Key where SA_BI_FOOTBALL_1210129.MatchesPlayers.player_out_id = DM_BI_FOOTBALL_1210129.DimPlayer.id
FactJogador_Jogo	playerKey	int	Fact	N/A	Staging Area, Data Mart	DimJogador	playerKey	int	Primary Key where SA_BI_FOOTBALL_1210129.MatchesPlayers.playerID = DM_BI_FOOTBALL_1210129.DimPlayer.id
FactJogador_Jogo	eventKey	int	Fact	N/A	Staging Area, Data Mart	DimEvent_Match	eventKey	int	Primary Key where SA_BI_FOOTBALL_1210129.MatchesPlayers.event_type = DM_BI_FOOTBALL_1210129.DimEvent_Match.event_type
DimEvento_Jogo	eventKey	int	Fact	N/A					Surrogate Key
DimEvento_Jogo	event_type	nvarchar (50)	Fact	1	Staging Area	Jogador_Jogo	event_type	nvarchar (30)	SELECT distinct (event_type) FROM MatchesPlayers
DimEvento_Jogo	bench	nvarchar (3)	Fact	1	Staging Area	Jogador_Jogo	bench	bit	

DimEvento_Jogo	lineup	nvarchar (3)	Fact	1	Staging Area	Jogador_Jogo	lineup	bit	
DimEvento_Jogo	home	nvarchar (3)	Fact	1	Staging Area	Jogador_Jogo	home	bit	
DimEvento_Jogo	away	nvarchar (3)	Fact	1	Staging Area	Jogador_Jogo	away	bit	
DimJogador	playerKey	int	Dimension	N/A					Surrogate Key
DimJogador	id	int	Dimension	1	Staging Area	Jogador	id	int	
DimJogador	full_name	nvarchar (160)	Dimension	1	Staging Area	Jogador	full_name	nvarchar (160)	
DimJogador	first_name	nvarchar (80)	Dimension	1	Staging Area	Jogador	first_name	nvarchar (80)	
DimJogador	last_name	nvarchar (120)	Dimension	1	Staging Area	Jogador	last_name	nvarchar (120)	
DimJogador	known_as	nvarchar (140)	Dimension	2	Staging Area	Jogador	known_as	nvarchar (140)	
DimJogador	height	int	Dimension	1	Staging Area	Jogador	height	int	
DimJogador	weight	int	Dimension	1	Staging Area	Jogador	weight	int	
DimJogador	position	nvarchar (40)	Dimension	2	Staging Area	Jogador	position	nvarchar (40)	
DimJogador	age	int	Dimension	1	Staging Area	Jogador	birthday	int	SELECT DATEDIFF(YEAR, CONVERT(DATE, DATEADD(SECOND, birthday, '19700101')),

									convert(date,getdate())) AS age FROM Players
DimJogador	countryKey	int	Dimension	N/A	Staging Area, Data Mart	DimCo untry	countryKe y	int	Foreign Key where SA_BI_FOOTBALL_1210129. Players.nationality = DM_BI_FOOTBALL_1210129 .DimCountry.country
DimJogador	birthdayKey	int	Dimension	N/A	Staging Area, Data Mart	DimDa te	yearKey	int	Foreign Key where CONVERT(DATE, DATEADD(SECOND, (SA_BI_FOOTBALL_1210129 .Players.birthday), '19700101')) = DM_BI_FOOTBALL_1210129 .DimDate.fullDate
DimJogador	effectiveDate	date	Dimension	N/A					
DimJogador	expiredDate	date	Dimension	N/A					
DimJogador	isCurrent	nvarchar (3)	Dimension	N/A					

4 Implementação da Solução

Neste Capítulo é detalhada toda a implementação da solução de acordo com o descrito no design. São apresentadas as configurações gerais do projeto e aprofundados todos os passos do processo ETL, abrangendo a extração, transformação (incluindo as operações de limpeza) e carregamento dos dados. Por último, é discriminado todo o método de criação dos cubos de processamento analítico online (OLAP) que permitirão a realização de análises.

4.1 Tecnologias Adotadas

Para a realização do projeto, foram eleitas tecnologias Microsoft, especificamente o SQL Server para criar o modelo de armazenamento de dados e SSIS para realizar a integração de dados. O SSIS, permite executar o processo de ETL (Extração, Transformação e Carregamento) e integrar facilmente com as ferramentas da Microsoft, incluindo o SQL Server. Para a análise aos dados é utilizada a ferramenta Power BI. A escolha das ferramentas é fundamentada com base nas características mencionadas no segundo Capítulo, bem como na familiaridade do autor com essas ferramentas.

4.2 Configurações Gerais

Com vista à implementação da solução, procedeu-se à aplicação de várias configurações gerais no SSIS, com o intuito de simplificar a sua execução. Estas englobam uma série de parâmetros, conexões e variáveis.

Nos parâmetros são definidos os nomes dos servidores, o ponto de origem da cache JogadoresJogos e o destino tanto da *Staging Area* (SA) como do Armazém de Dados (AD), a localização do projeto e os nomes das bases de dados a serem criadas. A maioria dos dados, uma vez extraídos via API da FootyStats, não requer qualquer parâmetro específico do projeto para essa finalidade. Na Figura 27 é possível visualizar esses mesmos parâmetros e respetivos valores.

Name	Data type	Value	Sensitive	Required	Description
DW_DBName	String	DW_BI_FOOTBALL_1210129	False	True	
MatchesPlayers_Cache	String	localhost	False	False	
PackagePath	String	C:\MEI\Tese_Final\1210129	False	True	
SA_DBName	String	SA_BI_FOOTBALL_1210129	False	True	
ServerName	String	localhost	False	True	

Figura 27 – Parâmetros do projeto

Ao executar todas as etapas da solução, estabelecem-se conexões de várias naturezas, incluindo ligação a ficheiros CSV como DimDate e DimTime_Match⁴, conexões com bases de dados e acesso à base de dados para cache (MachPlayers_Cache), a fim de criar tabelas que compõem as duas bases de dados. Na Figura 28, é apresentado o conjunto de todas as conexões estabelecidas para a concretização deste projeto.

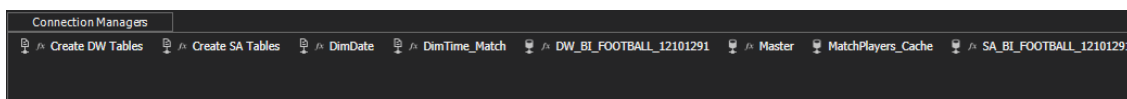


Figura 28 – Conexões do projeto

Para complementar a implementação das diversas componentes, foram criadas nove variáveis conforme demonstrado na Figura 29. As três primeiras representam os IDs das ligas, jogos e equipas, que são utilizadas durante a invocação dos *endpoints* para preencher as tabelas correspondentes, bem como outras tabelas relacionadas. Em seguida, tem-se a variável API_KEY, que contém a chave fornecida ao autor aquando da subscrição da API e que é parte

⁴ DimDate e DimTime_Match - As dimensões DimDate e DimTime_Match foram desenvolvidas utilizando o Microsoft Excel. Após a criação e formatação dos dados necessários, ambas as tabelas foram exportadas e armazenadas no formato CSV, o que facilitou sua importação e utilização no armazém de dados desenvolvido.

essencial de todas as chamadas aos *endpoints*. A variável *PaginationNumber* é usada em chamadas onde é necessário extrair dados de mais de uma página. Posteriormente, apresentam-se duas variáveis para a contagem de registos nas tabelas *DimDate* e *DimTime_Match*, respetivamente, de modo a identificar se estas estão vazias ou não. As restantes variáveis são criadas para facilitar a execução de *scripts* iterativas para a criação das diferentes tabelas em ambas as bases de dados.

Name	Scope	Data type	Value
LeagueID	Package	Object	System.Object
MatchID	Package	Object	System.Object
TeamID	Package	Object	System.Object
API_KEY	Package	String	1fb47b32852d0e348983f76648c6736c9d4c2ac08a2d8e8936a9e02ba29cd...
PaginationNumber	Package	Int32	1
DimDateNrOfRecords	Package	Int32	0
DimTimeNrOfRecords	Package	Int32	0
DWScriptFilename	Package	String	
SAScriptFilename	Package	String	

Figura 29 – Variáveis do projeto

4.3 Extração

Conforme mencionado na Secção 2.3.2, a extração representa a fase inicial do processo ETL, onde ocorre a transferência de dados dos sistemas de origem para uma área de *Staging* para posterior processamento.

Para a extração dos dados provenientes da API da FootyStats foi utilizado o SSIS Powerpack disponibilizado pela ZappySys[76]. O SSIS PowerPack, uma extensão desenvolvida especificamente para o SSIS da *Microsoft*, disponibiliza um conjunto abrangente de mais de 70 funcionalidades de fácil utilização, projetadas para automatizar tarefas, inclusive a integração de sistemas e manipulação de dados. Destaca-se especialmente na integração de dados por meio de APIs, com suporte para uma variedade de serviços, tais como “REST API, Azure Cloud, Amazon AWS Cloud, MongoDB, JSON, XML, CSV, Excel, Salesforce, Redshift, DynamoDB, Google API (incluindo Analytics e AdWords), SOAP/Web API, Facebook, Twitter, Zendesk, eBay, entre outros” [77].

Esta etapa iniciou-se pela criação da SA utilizando uma tarefa de execução de *script* SQL que combinada com o parâmetro acima ilustrado como *SA_DBName*, verifica se a mesma já foi criada, e caso contrário a base de dados em questão é gerada.

Seguidamente, utilizando um componente For Each Loop, são criadas iterativamente as diferentes tabelas utilizadas durante todo o processo de ETL. Para a criação das mesmas, o parâmetro denominado acima como PackagePath, identifica a localização da pasta do referido projeto onde dentro da mesma encontra-se a diretoria SQL_Scripts\StagingArea, na qual constam todas as *scripts* de geração (Anexo 4). Caso as tabelas já possuírem dados, as mesmas são primeiramente limpas, evitando assim a duplicação de dados. Na Figura 30, é possível visualizar essas duas etapas realizadas num projeto *Business Intelligence Integration Services* criado através da ferramenta Microsoft Visual Studio.

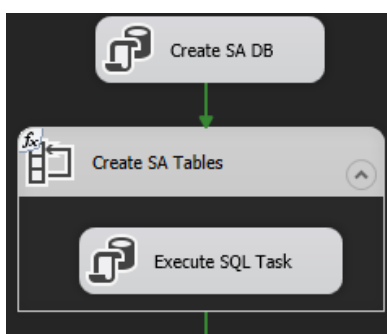


Figura 30 – Criação da SA

A Figura 31 apresenta a componente Data Flow Task, com detalhes adicionais na Figura 32, a qual desempenha um papel crucial na identificação dos vários IDs das principais competições ao longo das diversas épocas, nos dez países mais bem classificados no ranking da UEFA, conforme especificado na Secção 3.2. Além disso, esta componente é aplicada para carregar os dados relativos aos diferentes países, bem como para carregar um conjunto de dados em cache referentes à performance dos atletas nos vários jogos em que participaram. Esta abordagem foi adotada devido a uma restrição imposta pela FootyStats, que limita o número de chamadas por hora aos *endpoints*.

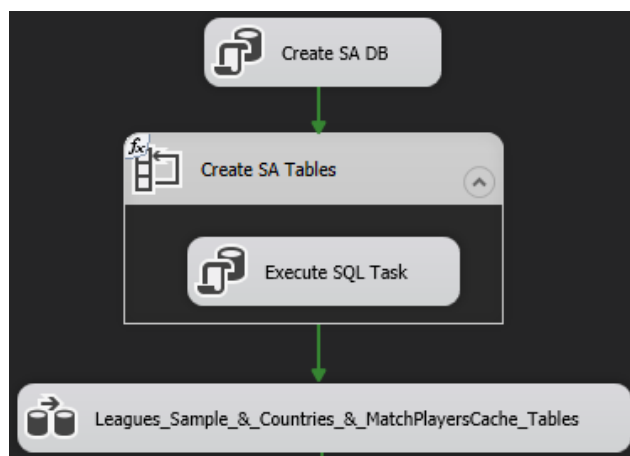


Figura 31 – Seleção de ligas e extração de primeiro conjunto de tabelas

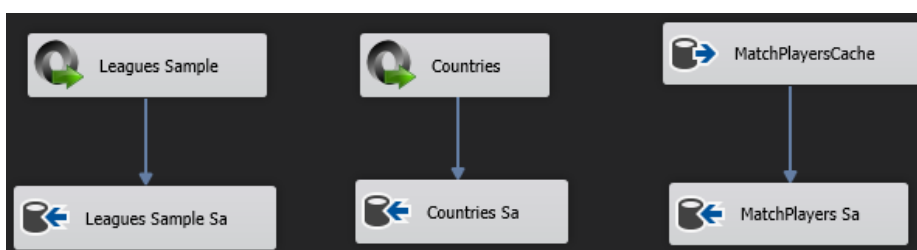


Figura 32 – Seleção de ligas e extração de primeiro conjunto de tabelas (Fluxo de Dados)

A componente JSON Source, disponibilizada no SSIS PowerPack, foi essencial para estabelecer a ligação entre os *endpoints* à SA. Esta componente específica é a indicada para a extração de dados em formato JSON da API da FootyStats, dentro do ambiente do SSIS. Com a aumento de APIs web e o crescente uso de JSON como formato de troca de dados, o JSON Source torna-se uma ferramenta indispensável para integrar dados provenientes dessas fontes. Além de possibilitar a extração direta de dados JSON, esta ferramenta permite também a consulta da estrutura dos dados JSON, a definição de filtros, a modelação e seleção dos dados a serem extraídos e ainda dispõe de ferramentas que facilitam a recolha de dados por paginação, entre outras funcionalidades. Todas estas características contribuem significativamente para simplificar o processo de incorporação de dados provenientes de API em formato JSON nos modelos dimensionais [78]. As Figuras 33, 34, 38 e 39 apresentam algumas das funcionalidades mencionadas, bem como dois *endpoints* distintos utilizados para a extração de dados.

Para a extração de amostra de competições foi utilizado o *endpoint* - https://api.football-data-api.com/league-list?key={{User::API_KEY}}&chosen_leagues_only=true. Neste endereço foram incluídos os parâmetros **key**, na qual é fornecida a chave obtida, e **chosen_leagues_only=true**, indicando assim que apenas as competições anteriormente selecionadas através da API da

FootyStats são invocadas. Ao nível do filtro aplicado aos dados foi utilizado \$.data[*].season[*], e foram selecionados apenas cinco atributos: *id*, *year*, *country*, *p_data_name* e *p_data_image*.

Na extração dos países, recorreu-se ao *endpoint* - https://api.football-data-api.com/league-list?key={{User::API_KEY}}. Neste endereço apenas se adicionou a key, sem aplicar qualquer filtro aos dados, mantendo-se \$.data[*], e selecionaram-se apenas cinco atributos: *id*, *iso*, *country*, *iso_number* e *name_pt*.

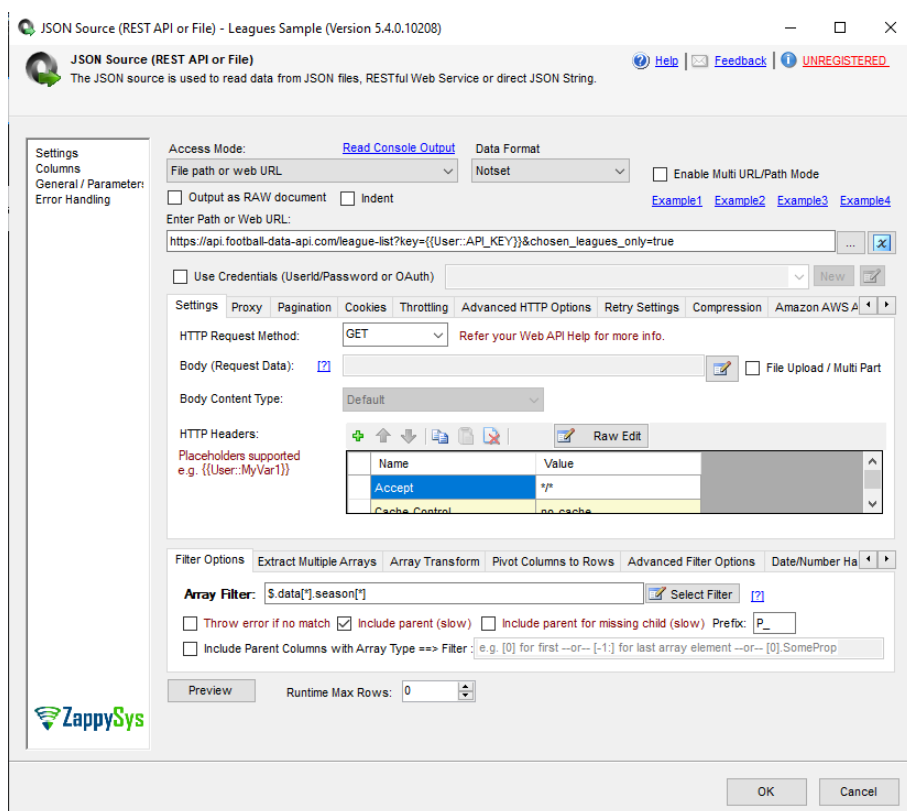


Figura 33 – Leagues Samples JSON Source (configurações do *endpoint*)

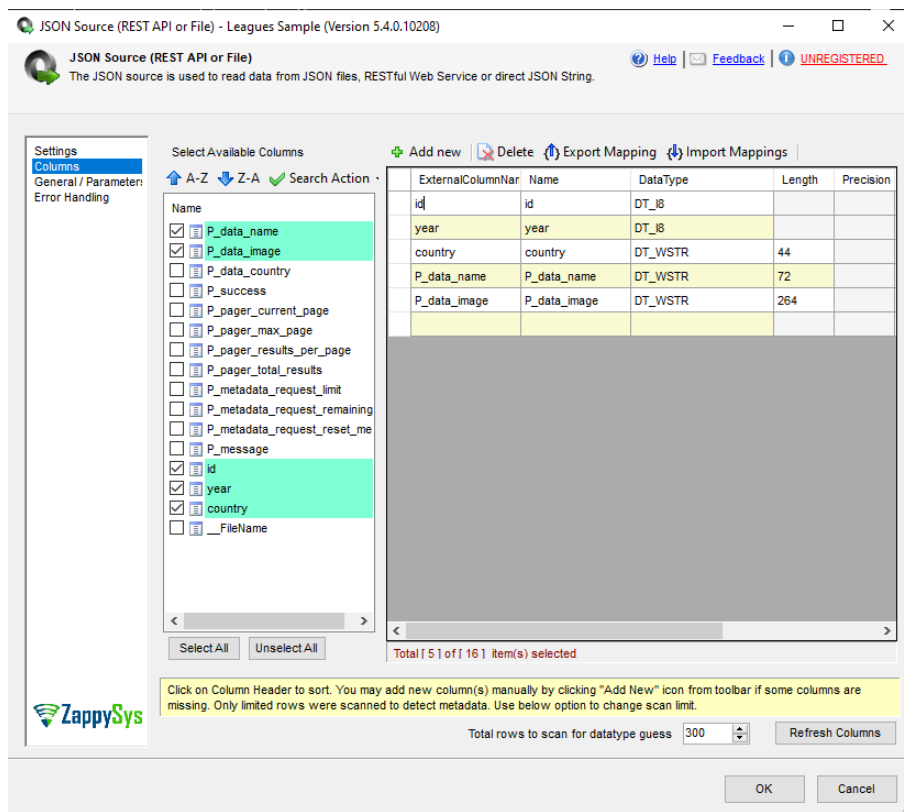
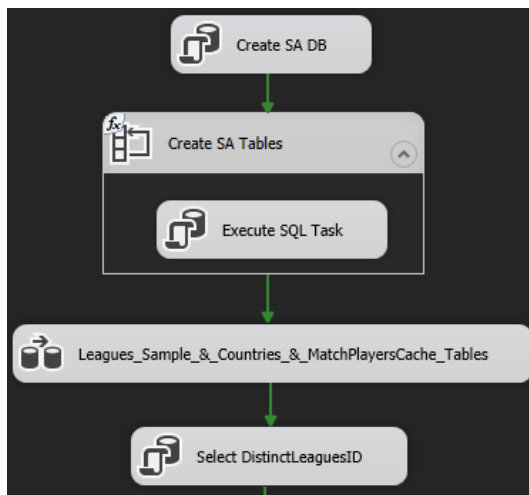


Figura 34 – Leagues Samples JSON Source (dados)

Após a identificação das competições de análise, é implementada uma tarefa de execução de *script* SQL para selecionar os diferentes IDs das competições. Esses IDs são posteriormente utilizados num componente For Each Loop, recorrendo-se da variável LeagueID, para chamar os diversos *endpoints* que requerem esse valor como parâmetro em seus endereços correspondentes. As Figuras 35 e 36 apresentam, respetivamente, a seleção dos IDs das competições e sua utilização.



```
SELECT DISTINCT Competition_ID
FROM Leagues_Sample
```

Figura 35 – Seleção de IDs distintos das ligas

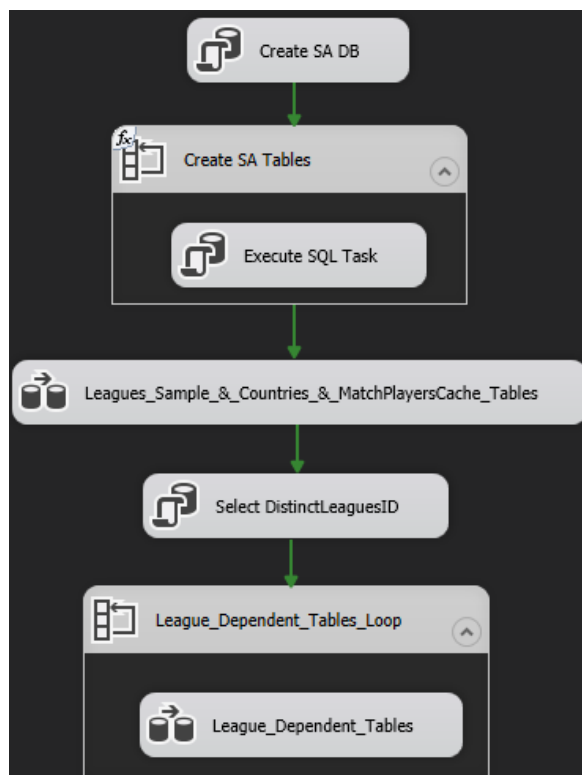


Figura 36 – Ciclo de extração de dados pendente de LeagueID

Na componente Data Flow Task incluída no For Each Loop, iterando os IDs das diferentes competições, realizou-se a extração de dados das tabelas Jogadores_Ligas, Equipas_Ligas, Jogos e Ligas. Além disso, foram criadas duas tabelas de referência, uma para equipas e outra para jogadores, conforme é possível averiguar na Figura 37. Estas tabelas foram criadas devido à inexistência de *endpoints* para a extração direta de valores únicos. Os dados disponíveis apenas se encontram organizados por competição, o que exigia a inserção de registos duplicados. Estas tabelas de referência foram fundamentais para a subsequente construção das respetivas tabelas com nomes idênticos e valores únicos. Esta etapa, apresentada na Figura 42, representa parte crucial da fase final do processo de Extração.

Na extração completa das competições, foi utilizado o *endpoint* - https://api.football-data-api.com/league-season?key={{User::API_KEY}}&season_id={{User::LeagueID}}. Neste endereço, os parâmetros *key* e *season_id* foram incluídos de forma iterativa. Este último identificado no passo anterior de identificação dos diferentes IDs das competições. Não foi aplicado qualquer tipo de filtro, mantendo-se `$.data[*]`, e foram selecionados diversos atributos, incluindo: *id*, *starting_year*, *ending_year* e *international_scale*.

No que concerne à recolha de dados sobre atletas e as diversas competições, e a tabela de referência de atletas, recorreu-se ao *endpoint* - https://api.football-data-api.com/league-players?key={{User::API_KEY}}&season_id={{User::LeagueID}}&page=<%pagination%>, conforme o disposto na Figura 38. Neste endereço, os parâmetros *key*, *season_id*, e *page* foram aplicados. Este último é particularmente usado para incluir uma iteração adicional na obtenção de dados das várias páginas, como já mencionado anteriormente. Para o efeito, foi utilizada a variável já apresentada, denominada de *PaginationNumber*, que é inicializada com o valor 1 e termina quando coincide com o valor indicado no atributo *\$.pager.max_page*, incluído no JSON do *endpoint* (Figura 39). Em ambas as chamadas, nenhum filtro foi aplicado, e na extração das competições, foram selecionados diversos atributos, tais como: *min_per_match*, *rank_in_league_top_attackers*, *min_per_goal_overall*. Já na referência dos jogadores, apenas os atributos de identificação dos atletas foram selecionados, entres os quais se destacam: *full_name*, *weight*, *height*, *position* e *nationality*.

No que diz respeito à extração de dados sobre ligas e equipas, bem como a tabela de referência de equipas, utilizou-se o *endpoint* - https://api.football-data-api.com/league-teams?key={{User::API_KEY}}&season_id={{User::LeagueID}}&include=stats. Este *endpoint* foi configurado com os parâmetros já mencionados *key*, *season_id*, adicionando o *include*, o qual dita a inclusão de estatísticas associadas. À semelhança dos restantes, não foi aplicado qualquer tipo de filtro, e foram selecionados diversos atributos como *seasonAVG_overall*, *winPercentage_overall* e *possessionAVG_overall* ao nível do preenchimento da tabela de ligas. Enquanto nas equipas foram incluídas apenas as propriedades identificativas, entre as quais se encontram o *id*, *name* e *founded*.

Por último, tanto na extração de jogos como nas jornadas partilham do mesmo *endpoint* - https://api.football-data-api.com/league-matches?key={{User::API_KEY}}&season_id={{User::LeagueID}}&max_per_page=1000. Utilizou-se os parâmetros *key*, *season_id* e excepcionalmente o *max_per_page*, sendo este último configurado com o valor 1000 para garantir o máximo de registos possíveis durante a chamada ao *endpoint*. Não foi aplicado qualquer tipo de filtro em ambas as chamadas e foram selecionados os atributos, *roundID*, *awayID*, *homeID* e *competition_id* ao nível das jornadas e diversos atributos quanto aos jogos, entre os quais incluem-se o *totalGoalCount*, *team_a_red_cards* e *team_b_penalty_goals*.



Figura 37 – Extração de dados pendentes de LeagueID

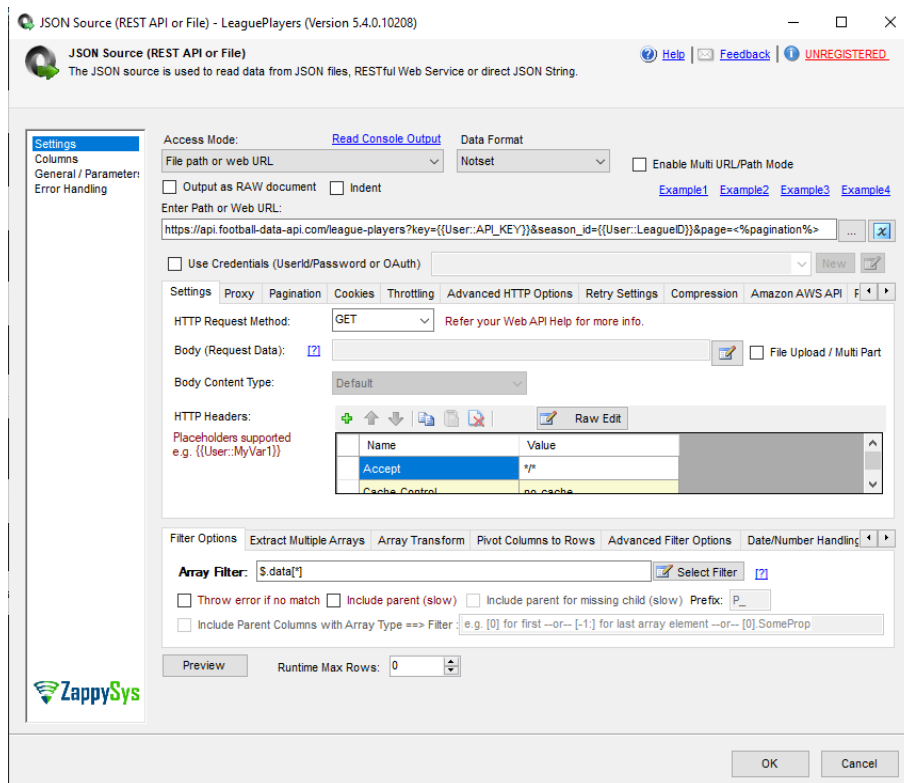


Figura 38 – LeaguePlayers JSON Source (configurações do endpoint)

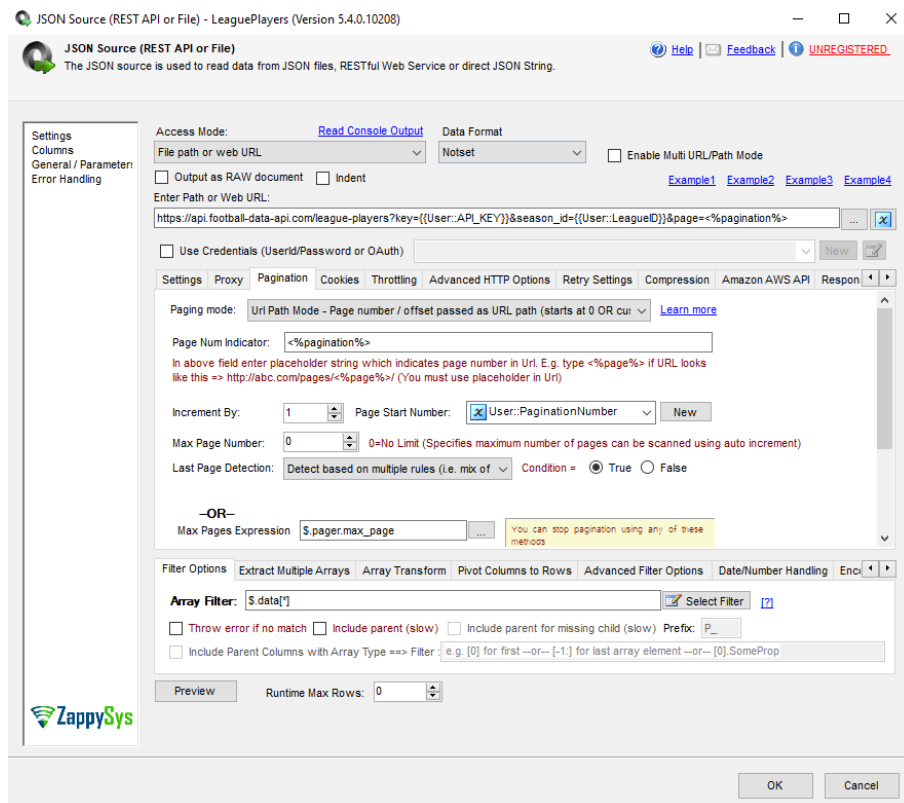


Figura 39 – LeaguePlayers JSON Source (paginação)

A fase de Extração é finalizada com o carregamento das tabelas de equipas, jogadores e continentes, sujeitas a um filtro de exclusão de valores duplicados (Sort), conforme demonstrado na Figura 42, juntamente com a identificação única dos IDs das equipas e jogos, realizada através da execução de um *script* SQL (Figura 41). Estes identificadores são posteriormente utilizados num componente For Each Loop, recorrendo-se à variável TeamID e MatchID, respetivamente, para as chamadas aos *endpoints* que requerem esses parâmetros. Conforme já mencionado, a tabela Jogadores_Jogos foi abastecida maioritariamente com dados armazenados em cache local, sendo nesta etapa apenas inseridos registos correspondentes ao diferencial da mesma. Através da Figura 40, é possível visualizar todo o processo completo da criação da SA.

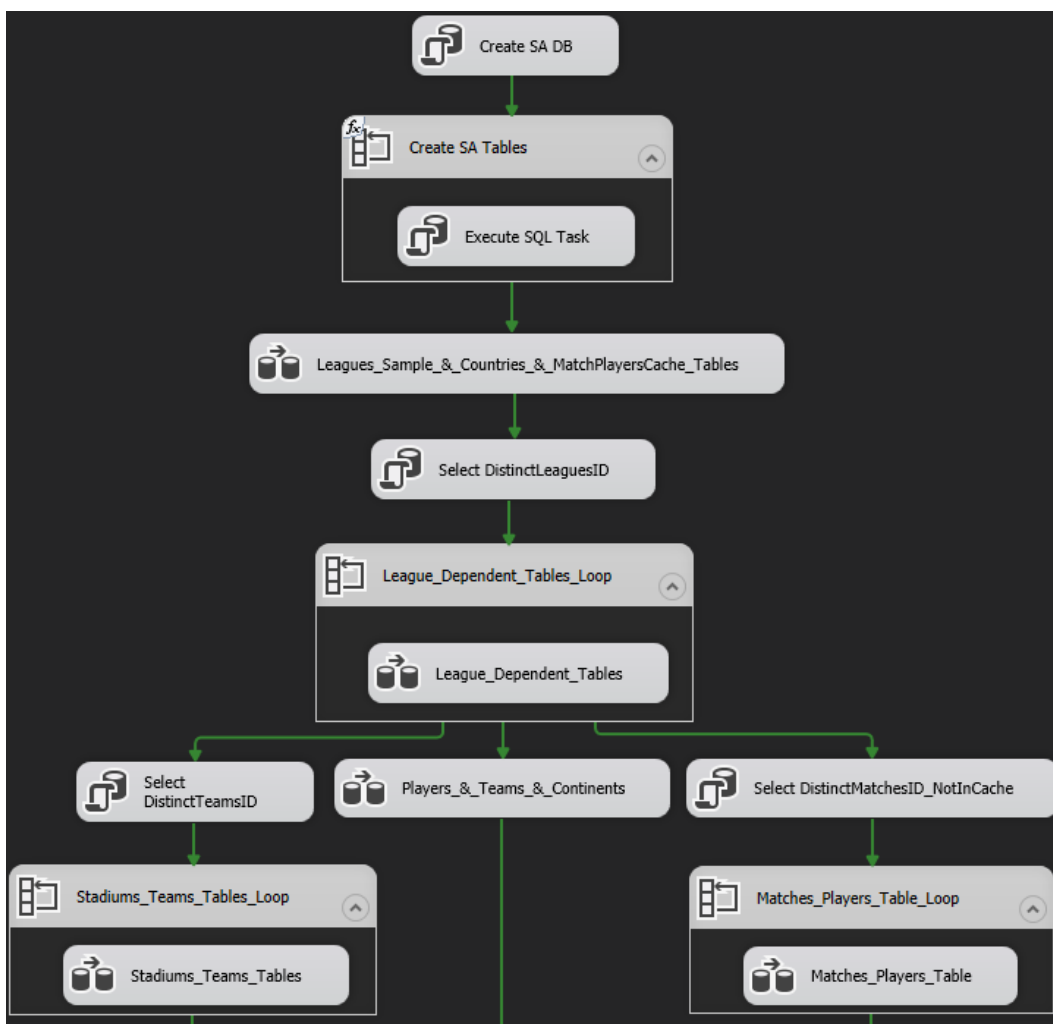


Figura 40 – Processo de criação da SA

```

SELECT DISTINCT homeID
FROM    Matches

SELECT DISTINCT ID
FROM    Matches
where id not in (SELECT matchID FROM MatchesPlayers)
and status like 'Complete'

```

Figura 41 – Código de seleção distinta das equipas e jogos completos

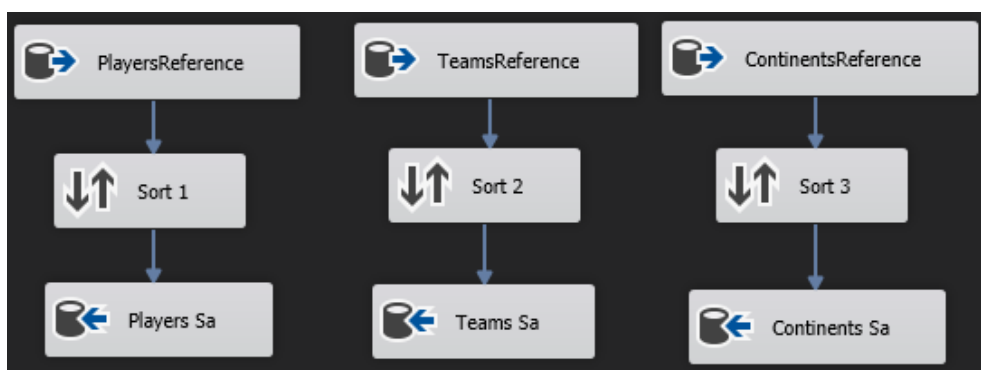


Figura 42 – Extração de dados de jogadores, equipas e continentes

Na obtenção de dados sobre estádios, recorreu-se ao *endpoint* - https://api.football-data-api.com/team?key={{User::API_KEY}}&team_id={{User::TeamID}}. Para este *endpoint* foram utilizados os parametros *key* e *team_id*. Não houve aplicação de filtros, tendo sido selecionados os atributos, *stadium_address*, *stadium_name* e *team_id*. Este último referenciado para identificar os estádios únicos pertencentes às equipas correspondentes.

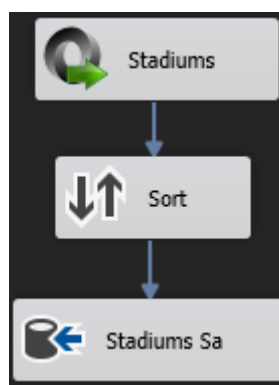


Figura 43 – Extração de dados de estádios

Na extração do dados em cache como do diferencial relativos à performance dos jogadores nas diversas partidas de futebol que participaram foi utilizado o *endpoint* - <https://api.football->

data-api.com/match?key={{User::API_KEY}}&match_id={{User::MatchID}}, no qual foram utilizados os parâmetros key e team_id. O formato JSON disponibilizado por este endereço apresenta a particularidade de dispor a informação em diversos *arrays* aninhados, conforme se pode observar na Figura 44.

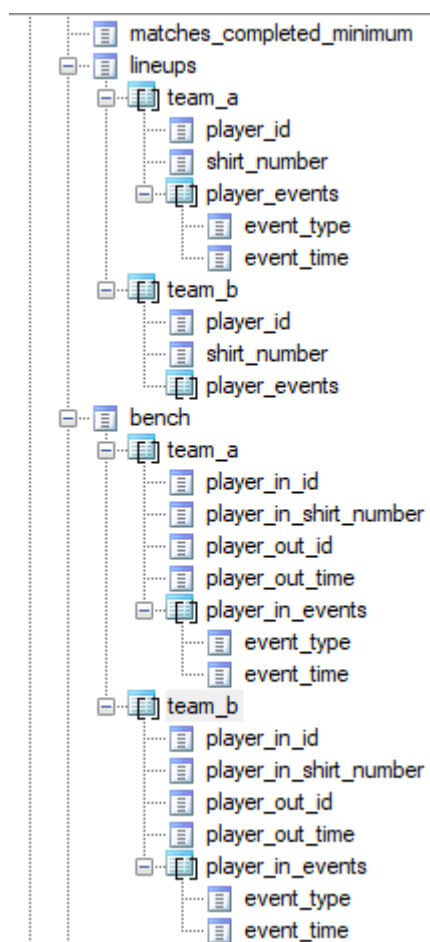


Figura 44 – Formato JSON do endereço para extração de dados de jogadores_jogos

Para viabilizar a organização dos dados numa tabela, tirou-se proveito da funcionalidade de filtro disponível pela JSON Source. Os jogadores foram divididos em oito grupos distintos: aqueles que jogam em casa no onze titular com alguma ação em jogo (`$.data.lineups.team_a[*].player_events[*]`), os que jogam fora no onze titular com alguma ação em jogo (`$.data.lineups.team_b[*].player_events[*]`), os que jogam em casa no banco de suplentes com alguma ação em jogo (`$.data.bench.team_a[*].player_in_events[*]`), os que jogam fora no banco de suplentes com alguma ação em jogo (`$.data.bench.team_b[*].player_in_events[*]`), os que jogam em casa no onze titular sem ação em jogo (`$.data.lineups.team_a[*]`), os que jogam fora no onze titular sem ação em jogo

(\$.data.lineups.team_b[*]), os que jogam em casa no banco de suplentes sem ação em jogo (\$.data.bench.team_a[*]) e os que jogam fora no banco de suplentes sem ação em jogo (\$.data.bench.team_b[*]). Para identificar os jogos sem eventos, foi utilizado um Condition Split Transformer com a expressão LEN (player_events) == 2, onde a estrutura do *array* vazio é representada por [] indicando a ausência de eventos. Para facilitar a identificação dos grupos mencionados anteriormente, foram utilizados campos booleanos, como *bench*, *lineup*, *home* e *away*, preenchidos de acordo com o grupo ao qual pertencem. A tabela é então preenchida por meio de diversas uniões entre esses registros dos grupos distintos. A Figura 45 ilustra toda essa metodologia implementada.

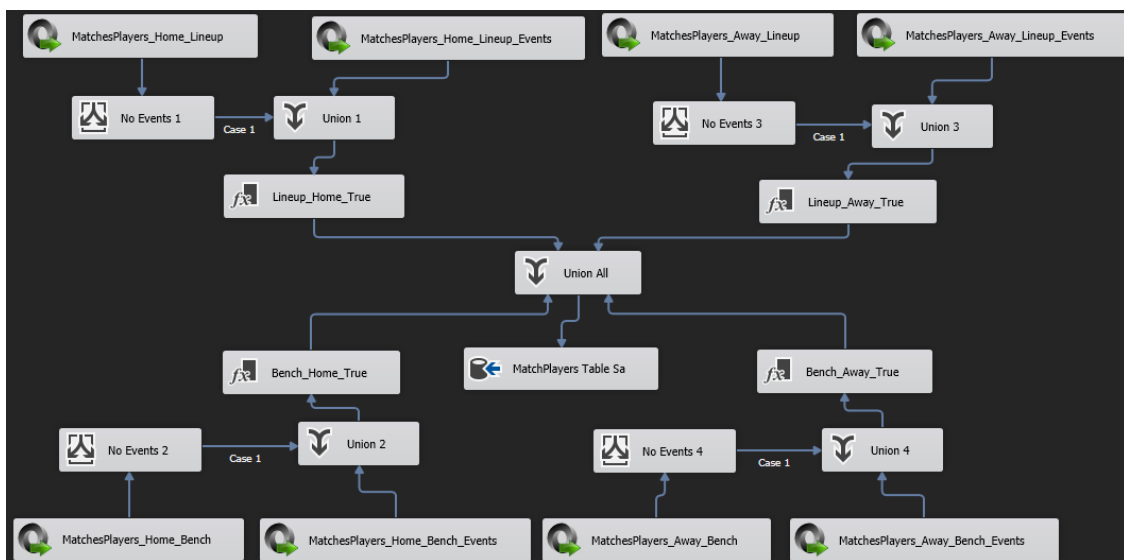


Figura 45 – Extração de diferencial de dados de performance de jogadores em jogos

4.4 Transformação e Carregamento

A segunda e terceira fases do processo ETL são a transformação e o carregamento, respectivamente. Estas etapas estão interligadas, pois logo após a transformação dos dados, procede-se ao carregamento das dimensões ou tabelas de factos correspondentes, imediatamente a seguir.

À semelhança da criação da SA e as respectivas tabelas, é utilizada uma tarefa de execução de *script SQL*, que, através do parâmetro do projeto DW_DBName, verifica a existência do AD. Se não estiver criado, a base de dados é então gerada. Em seguida, utilizando um componente For Each Loop, as diferentes tabelas que compõem o AD são criadas de forma iterativa. Para isso,

os *scripts* de geração estão localizados na pasta SQL_Scripts\DataWarehouse, dentro da estrutura do projeto. A Figura 46 ilustra o processo de criação do AD e suas tabelas correspondentes.

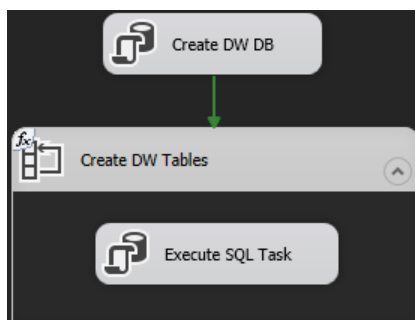


Figura 46 – Criação do AD e respetivas tabelas

Após a criação das tabelas precede-se aos carregamentos das diferentes dimensões e tabelas de facto. As primeiras dimensões a serem carregadas são a DimDate e a DimTime_Match, recorrendo-se a um Sequence Container. Estes carregamentos são executados através das conexões estabelecidas com ficheiros .csv correspondentes, com a mesma nomenclatura, como representado na Figura 28. Antes de efetuar a importação dos dados, verifica-se se já existem registos, utilizando as variáveis DimDateNrOfRecords e DimTimeNrOfRecords, respetivamente. Caso não existam, os dados são importados a partir dos ficheiros correspondentes (Figura 47). Se os dados já existirem, o componente é ignorado. O fluxo de dados no carregamento de ambas as tabelas é apresentado na Figura 48.

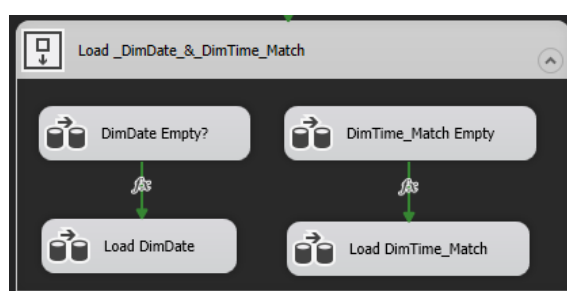


Figura 47 – Carregamento da DimDate e DimTime_Match

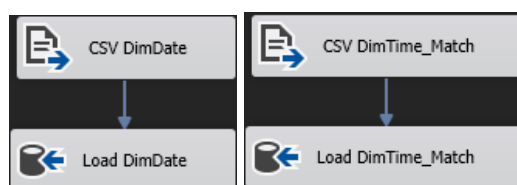


Figura 48 – Processo de carregamento da DimDate e DimYear DimTime_Match

Assim como as tabelas DimDate e DimTime_Match foram criadas, as dimensões de negócio são igualmente carregadas por meio de um Sequence Container à exceção da DimCountry (Figura 49).

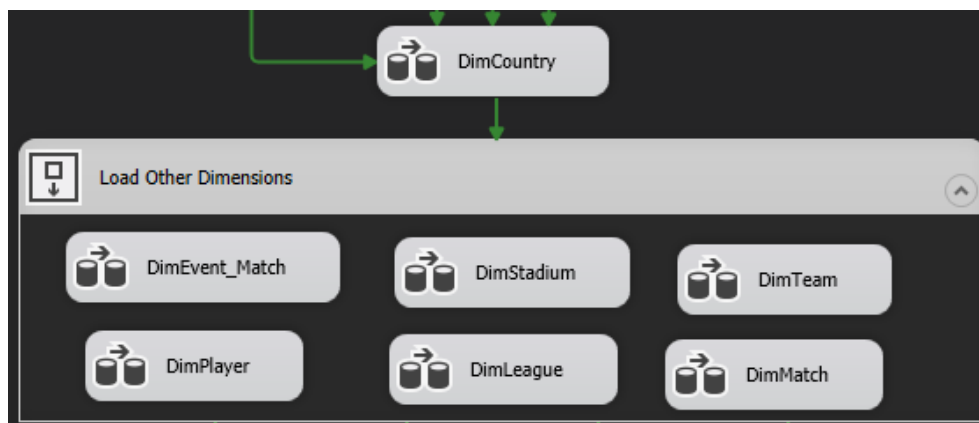


Figura 49 – Carregamento de restantes Dimensões

O carregamento da DimCountry inicia-se com a recolha da tabela Countries, onde os dados são transformados usando o campo *iso* para determinar os continentes e acrónimos correspondentes. Caso o campo *iso* esteja vazio ou não corresponda a nenhum continente conhecido, os países são classificados como continente "Desconhecido". Em seguida, é realizado um left join com os dados da tabela Continents. Os registos identificados como "Desconhecido" são então encaminhados para uma tabela CountriesDQP, que reúne todas as inconsistências de dados relativos aos países. Dada a dimensão dos dados analisados nesta dissertação, os *Data Quality Problems* (DQP) não serão objeto de análise ou tratamento, sendo excluídos por agora, podendo ser considerados em trabalho futuro. Como os dados presentes na DimCountry são apenas Slow Changing Dimension (SCD) do tipo 1, o processo de carregamento é finalizado apenas com uma componente de atualização, conforme o exibido na Figura 50.

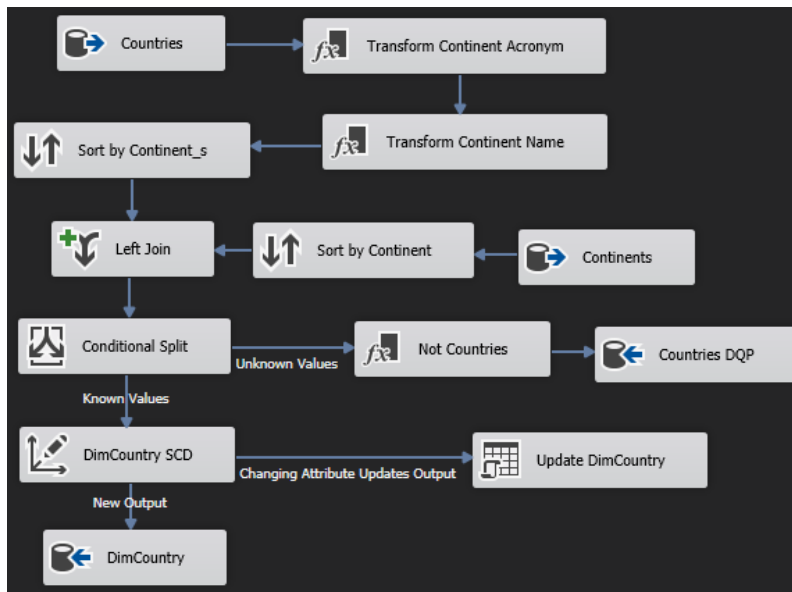


Figura 50 – Processo de carregamento da DimCountry

O processo de carregamento da Junk Dimension DimEvent_Match, por sua vez, começa com a recolha dos dados presentes na tabela Match_Players, onde são ordenados e filtrados com base nas combinações possíveis dos campos *event_type*, *bench*, *lineup*, *home*, *away*, sem registos duplicados. Para cada um destes quatro últimos campos, é realizado um Lookup nas tabelas, BenchFlag_Lookup, LineupFlag_Lookup, AwayFlag_Lookup, HomeFlag_Lookup, onde os valores 0 e 1 são substituídos por “No” e “Yes”, respetivamente. Caso não haja correspondência, os registos são então redirecionados para a tabela Event_MatchDQP. À semelhança da DimCountry, este processo é também finalizado através de uma componente de atualização, uma vez que os campos são apenas SCD do tipo 1. A Figura 51 apresenta o referido processo de carregamento de dados.

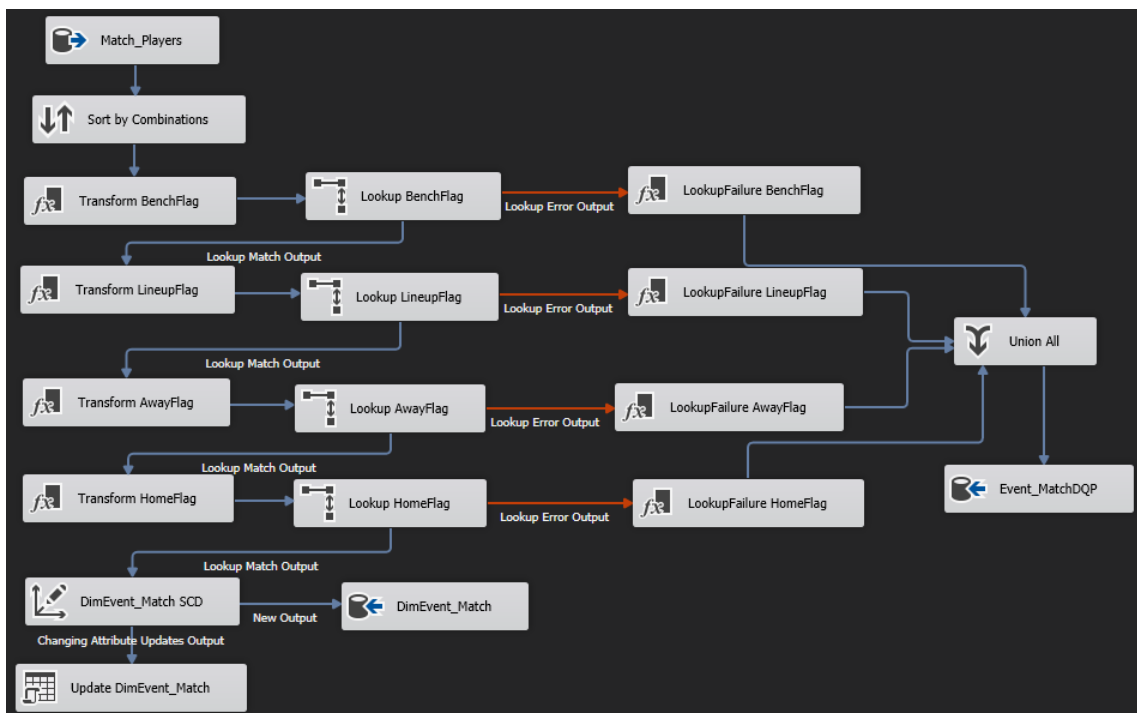


Figura 51 – Processo de carregamento da DimEvent_Match

A DimStadium é alimentada com os dados provenientes da tabela Stadiums. Inicialmente, procede-se à correção dos endereços dos estádios, eliminando inconsistências como ", Jamor, Oeiras", para "Jamor, Oeiras". As outras inconsistências que não têm solução são encaminhadas para StadiumsDQQ, enquanto aquelas que seguem o curso normal são filtradas de forma a excluir-se registos duplicados. Como existem dados em que se pretende preservar histórico, o carregamento é concluído com a atualização dos campos *expiredDate* e *effectiveDate*, bem como o campo *IsCurrent* (Figura 52). Para isso, utiliza-se a variável de sistema, (@[System::ContainerStartTime]).

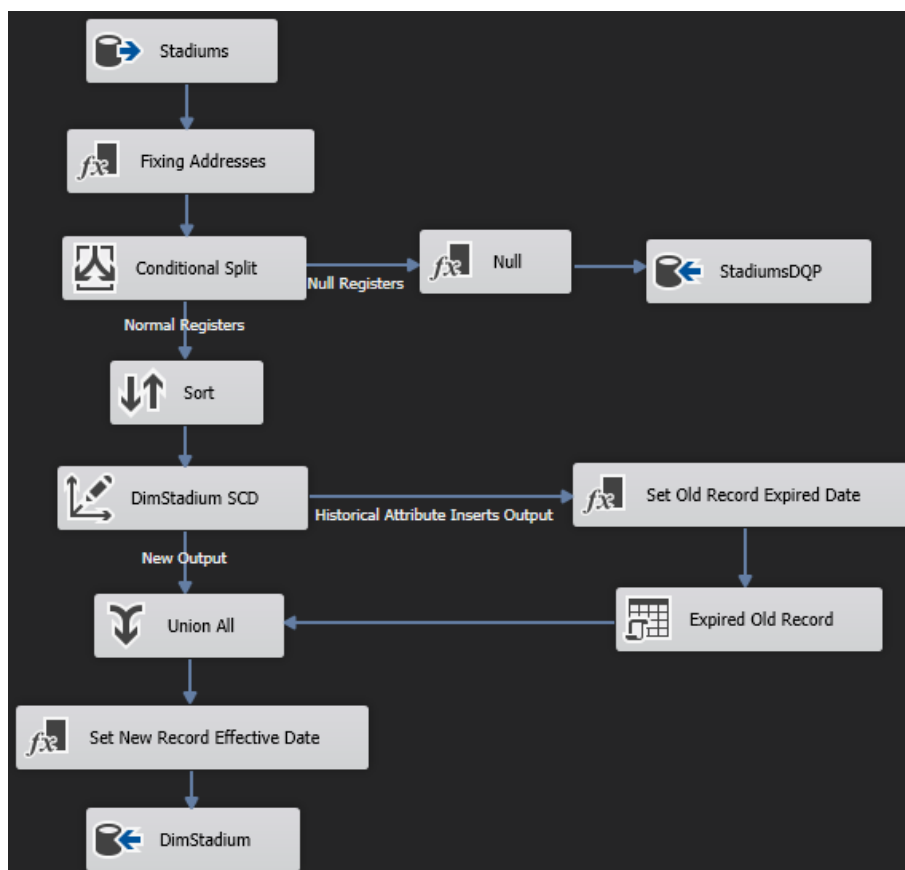


Figura 52 – Processo de carregamento da DimStadium

O carregamento da DimPlayer tem início com a obtenção da tabela Players, na qual são excluídos os registos duplicados com base no campo *id*. Posteriormente, são aplicadas transformações ao campo *birthday*. Primeiramente, os valores 0 são substituídos por nulo, evitando assim datas de nascimento incorretas e, conseqüentemente, idades inconsistentes. Em seguida, o formato UNIX é convertido para o tipo Date. Em etapas subsequentes, são realizados Lookups às tabelas DimCountry e DimDate, onde os valores de *nationality* e *birthday* transformado são cruzados com *country* e *fulldate*, recolhendo-se as respetivas chaves. Seguidamente, os valores -1 nos campos *weight* e *height* são substituídos por nulo e as idades de cada jogador são calculadas subtraindo-se a data de nascimento (*birthday*) da data atual (Figura 53). Assim como a DimStadium, na DimPlayer existem campos SCD tipo 2, o que implica a atualização dos campos *expiredDate* e *effectiveDate*, bem como do campo *IsCurrent*, concluindo assim o processo.

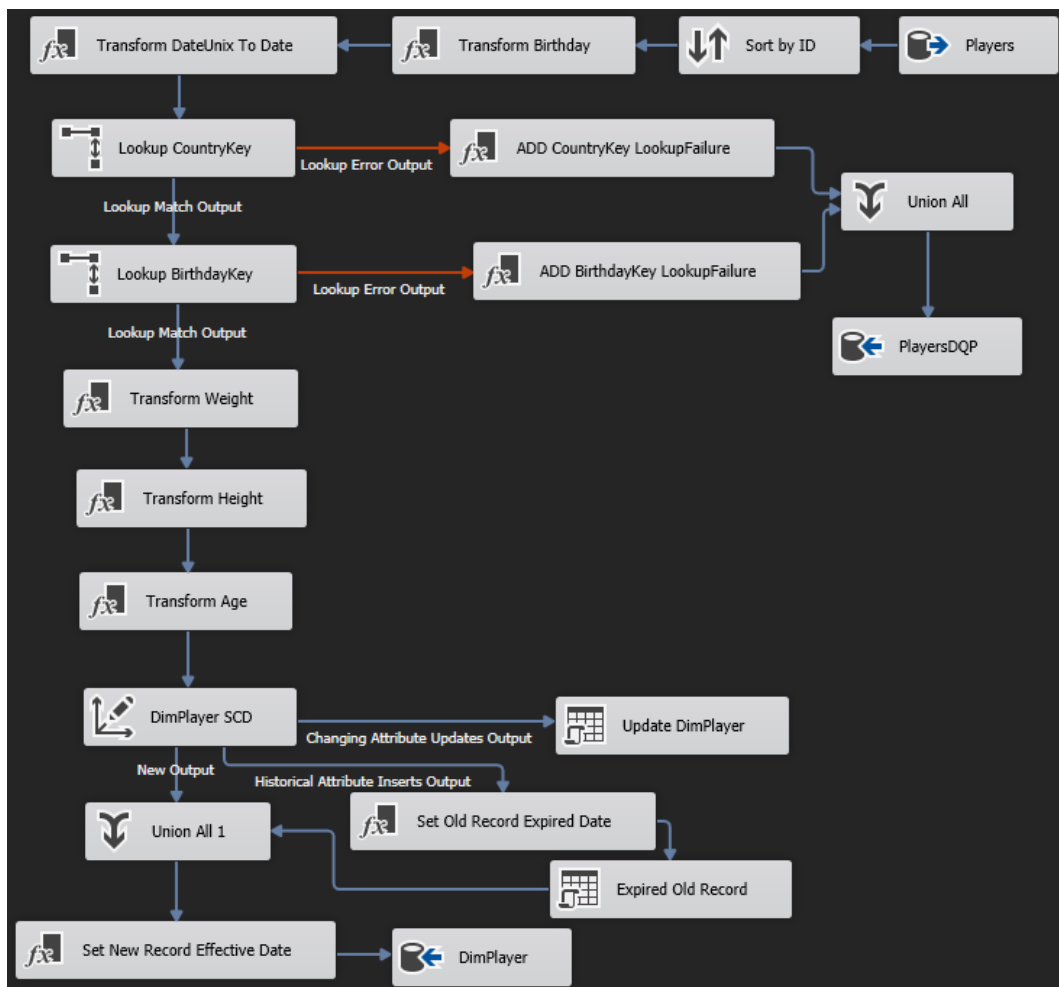


Figura 53 – Processo de carregamento da DimPlayer

Os dados contidos na dimensão DimLeague são provenientes da tabela Leagues. Primeiramente, são extraídos os anos do campo *season*, que apresenta dois anos separados por uma barra (Ex: 2023/2024). Em seguida, esses anos são cruzados com o campo *year* na DimDate através de um Lookup, resultando na obtenção da chave correspondente. Os registros que não possuem correspondência são então dirigidos para a tabela LeaguesDQP. Tal como anteriormente, na DimLeague existem atributos que se pretende registrar a evolução histórica. Assim, este processo de carregamento é finalizado utilizando uma componente SCD, que antecede a atualização dos campos *expiredDate* e *effectiveDate* e *IsCurrent*. A Figura 54 apresenta todo o referido processo de carregamento da dimensão DimLeague.



Figura 54 – Processo de carregamento da DimLeague

A DimTeam, a última dimensão a ser carregada, os registos duplicados são excluídos com base no campo *ID* em primeiro lugar. Seguidamente, somente o primeiro ano é extraído para evitar inconsistências como “1907 / 2005 / 2012 / 2013”. Para cada ano é realizado um Lookup à DimDate para obter a chave correspondente. Caso não haja correspondência, os registos são direcionados para a tabela TeamsDQP. As equipas que possuem dados coerentes têm seus anos de existência apurados. O processo é finalizado com uma componente de atualização dos registos, já que se trata apenas de campos SCD tipo 1 (Figura 55).

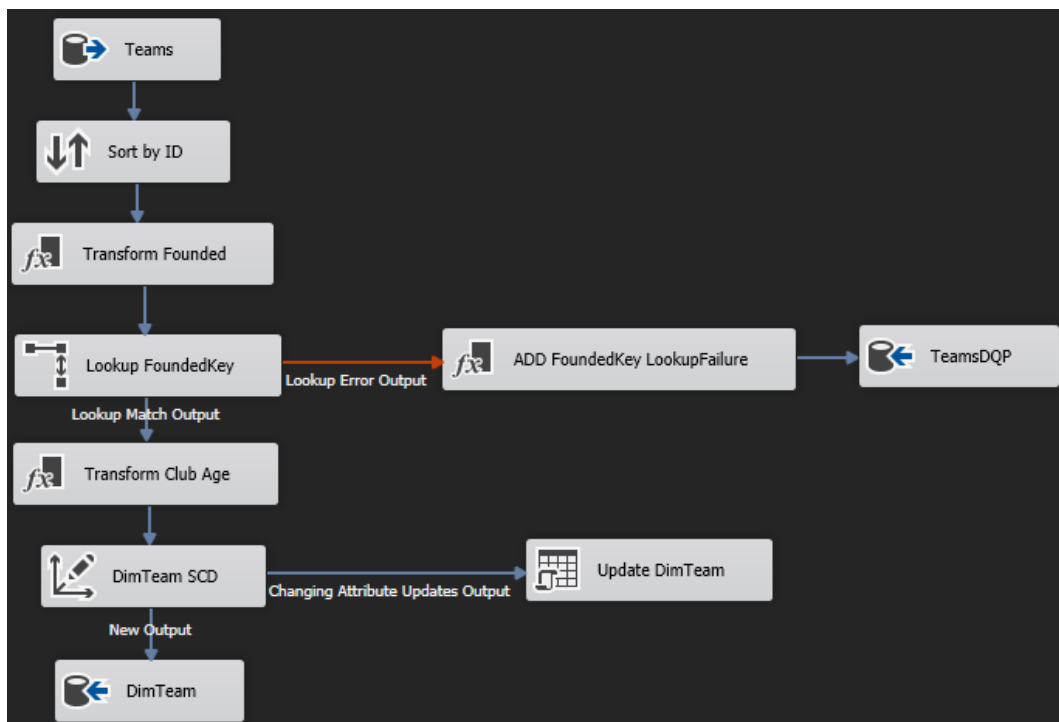


Figura 55 – Processo de carregamento da DimTeam

A DimMatch, como a última dimensão a ser carregada, verifica a existência de registos utilizando um Lookup através do ID. Na ausência de registos, procede-se ao carregamento dos dados. Este processo de carregamento dos identificadores das partidas nesta dimensão está ilustrado na Figura 56.

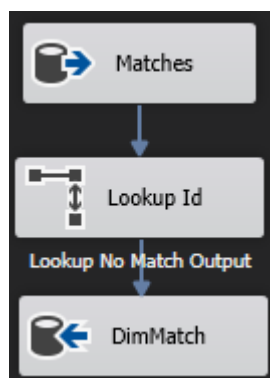


Figura 56 – Processo de carregamento da DimMatch

Com todas as dimensões preenchidas e os respetivos dados prontos para a sua utilização, procede-se ao carregamento das tabelas de factos: FactLeague_Player (Figura 57), FactLeague_Team (Figura 59), FactMatch e FactMatch_Player (Figura 61).

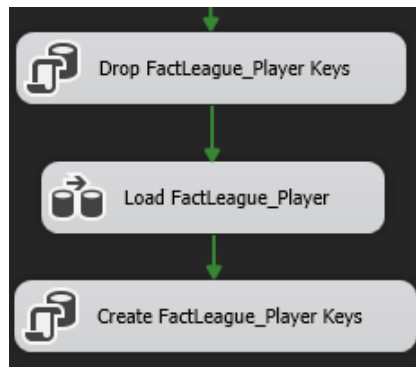


Figura 57 – Carregamento da FactLeague_Player

O carregamento da tabela de factos FactLeague_Player, é realizado através do acesso aos dados presentes na tabela LeaguePlayers. Durante este processo, são realizadas diversos Lookups nas tabelas DimPlayer, DimLeague, DimTeam e DimDate para obter-se as respetivas chaves. Registos que não possuam correspondência são redirecionados para a tabela FactLeague_PlayerDQP, com a exceção do Lookup Tem_2Key, cujo objetivo é apenas obter a chave para registos com valores. Adicionalmente, são realizadas duas transformações: a primeira converte o campo *last_match_timestamp* de UNIX para Date, enquanto a segunda substitui os valores -1 no campo *club_team_2_id* por nulo. O processo é concluído com a verificação do facto, atualizando-o se já existir ou carregando-o pela primeira vez, caso contrário, como ilustrado na Figura 58.

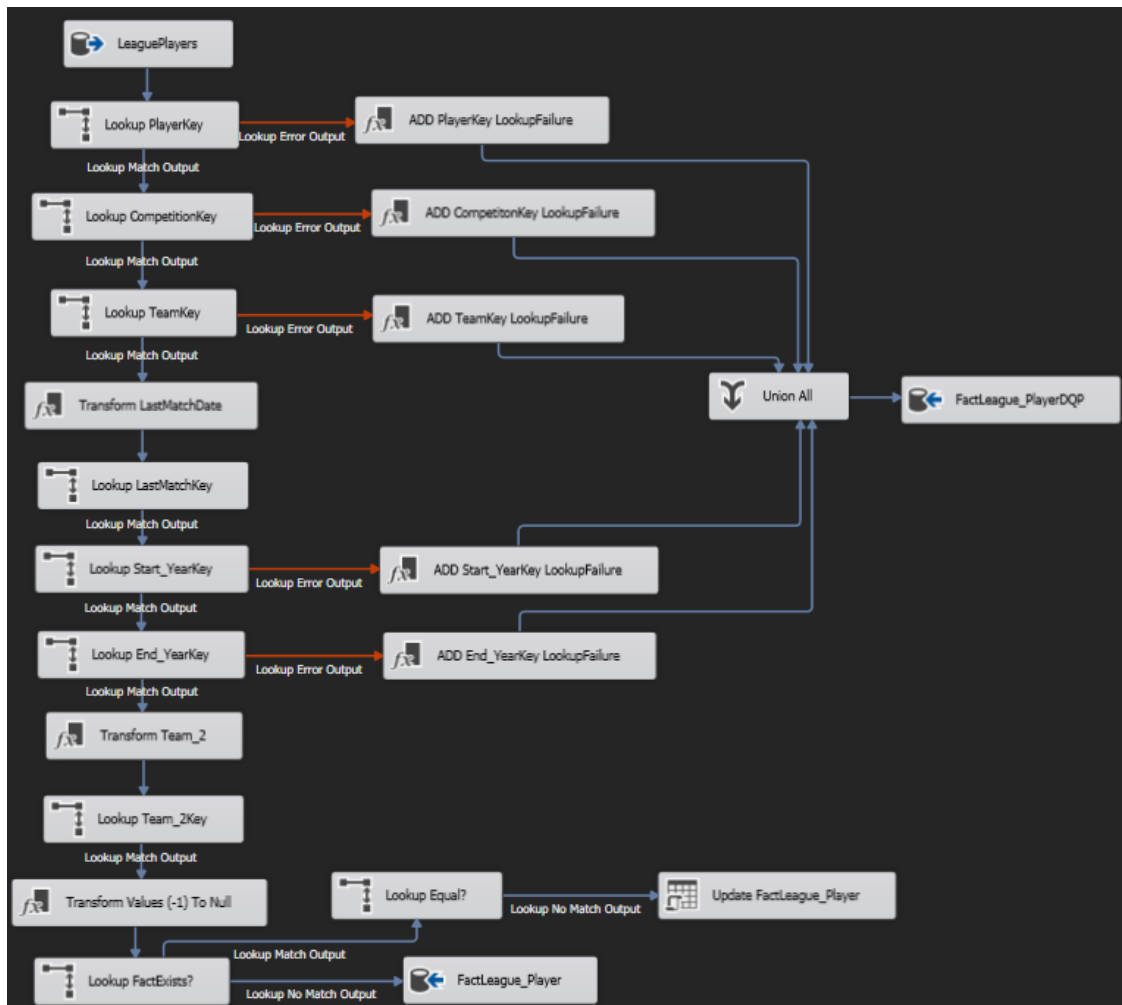


Figura 58 – Processo de carregamento da FactLeague_Player

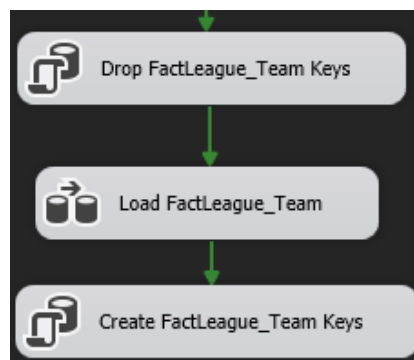


Figura 59 – Carregamento de FactLeague_Team

O carregamento da tabela de factos FactLeague_Team, por sua vez, é realizado recolhendo dados da tabela LeagueTeams. Nesse procedimento, são realizados diversos Lookups nas tabelas DimTeam, DimLeague e DimDate, recolhendo-se as respetivas chaves. Caso não haja

correspondência, os registos são encaminhados para a tabela FactLeague_TeamDQP. Assim como na inserção da DimLeague, os anos são extraídos separadamente antes dos respetivos Lookups. A Figura 60 exibe todo o procedimento de carregamento da FactLeague_Team.

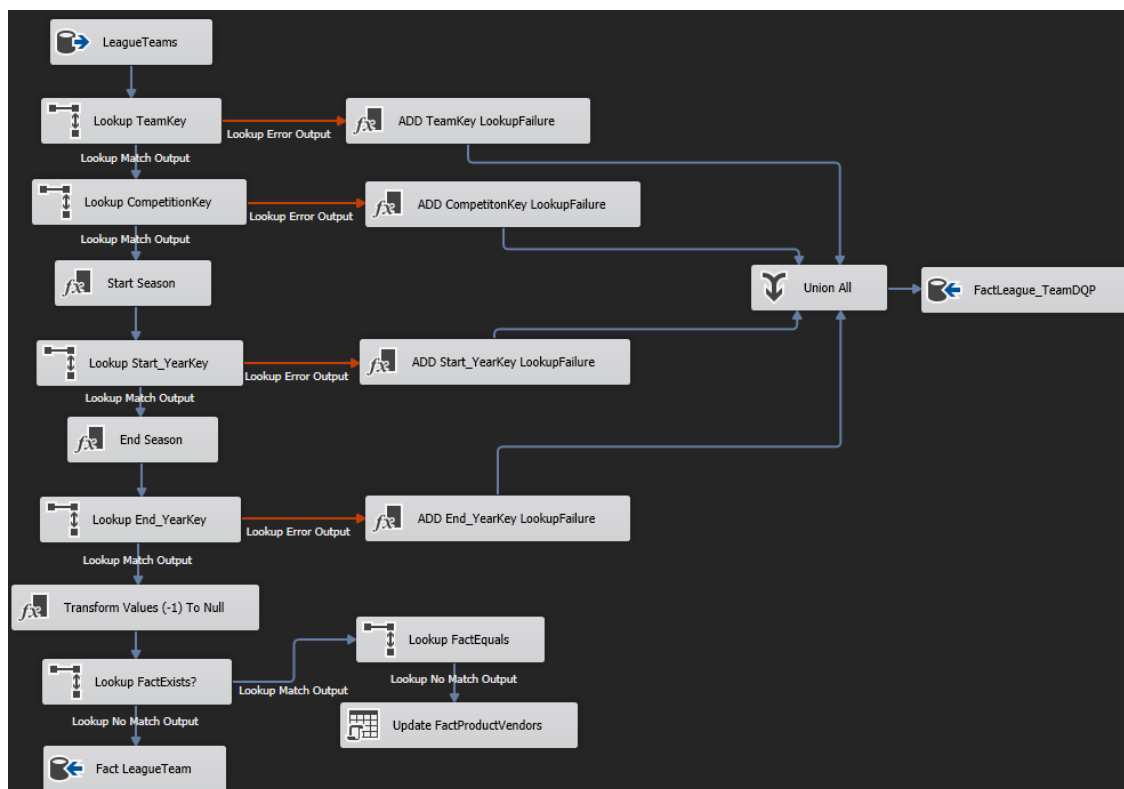


Figura 60 – Processo de carregamento de FactLeague_Team



Figura 61 – Carregamento de FactMatch e FactMatch_Player

A tabela de factos FactMatch, é carregada através dos dados vindouros das tabelas Matches e Rounds unidas através de um Left Join. Seguidamente, são realizados diversos Lookups na tabela DimTeam, para apurar-se as chaves das equipas visitada e visitante, bem como nas tabelas DimLeague, DimDate, DimMatch e DimStadium. Caso não haja correspondência, os

registos são encaminhados para a tabela FactMatchDQP. Durante este processo, ilustrado na Figura 62, o formato UNIX do campo *date_unix* é convertido para DATE.

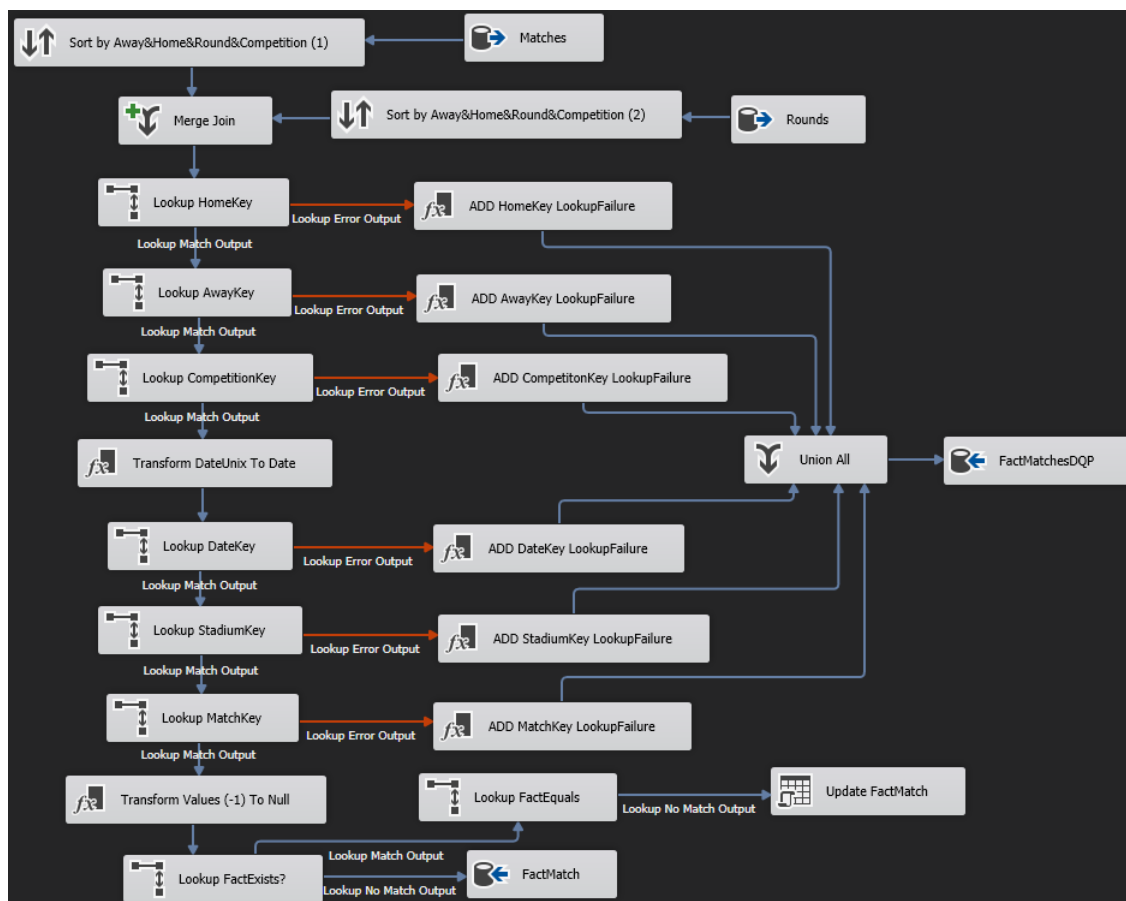


Figura 62 – Processo de carregamento de FactMatch

O processo ETL é concluído ao carregar a tabela de factos sem factos (em inglês: *factless fact table*) FactMatch_Player. Assim como nas outras tabelas de factos, são realizados Lookups em todas as dimensões, exceto na DimCountry. As correspondências sem sucesso são redirecionadas para a tabela FactMatch_PlayerDQP, com exceção do Lookup Player_out_TimeKey, onde apenas a chave dos registos com valor é identificada. Durante este procedimento, apresentado na Figura 63, os valores 0 e -1 nos campos *shirt_number*, *player_out_id* e *player_out_time* são substituídos por nulo.



Figura 63 – Processo de carregamento de FactMatch_Player

4.5 Cubo OLAP

Os cubos OLAP representam estruturas multidimensionais que agilizam a análise de dados. Para conduzir tais análises, são empregues dados diversos, principalmente oriundos de Armazéns de Dados (AD) ou Data Mart (DM). A partir de um AD ou DM, é possível criar vários cubos,

utilizando diferentes dimensões e tabelas de fatos, para analisar os dados sob várias perspectivas. Essa prática é motivada por diversos motivos entre os quais estão: o desempenho, a segurança ou a conformidade dos dados [23], [30].

No âmbito deste projeto, foi desenvolvido um cubo dedicado à análise de dados sobre o desempenho das equipas e dos atletas nas suas competições, bem como aos eventos ocorridos tanto a nível das equipas como dos atletas durante as partidas. A Figura 64 ilustra o Cubo OLAP criado, com as tabelas de factos e de dimensões correspondentes.

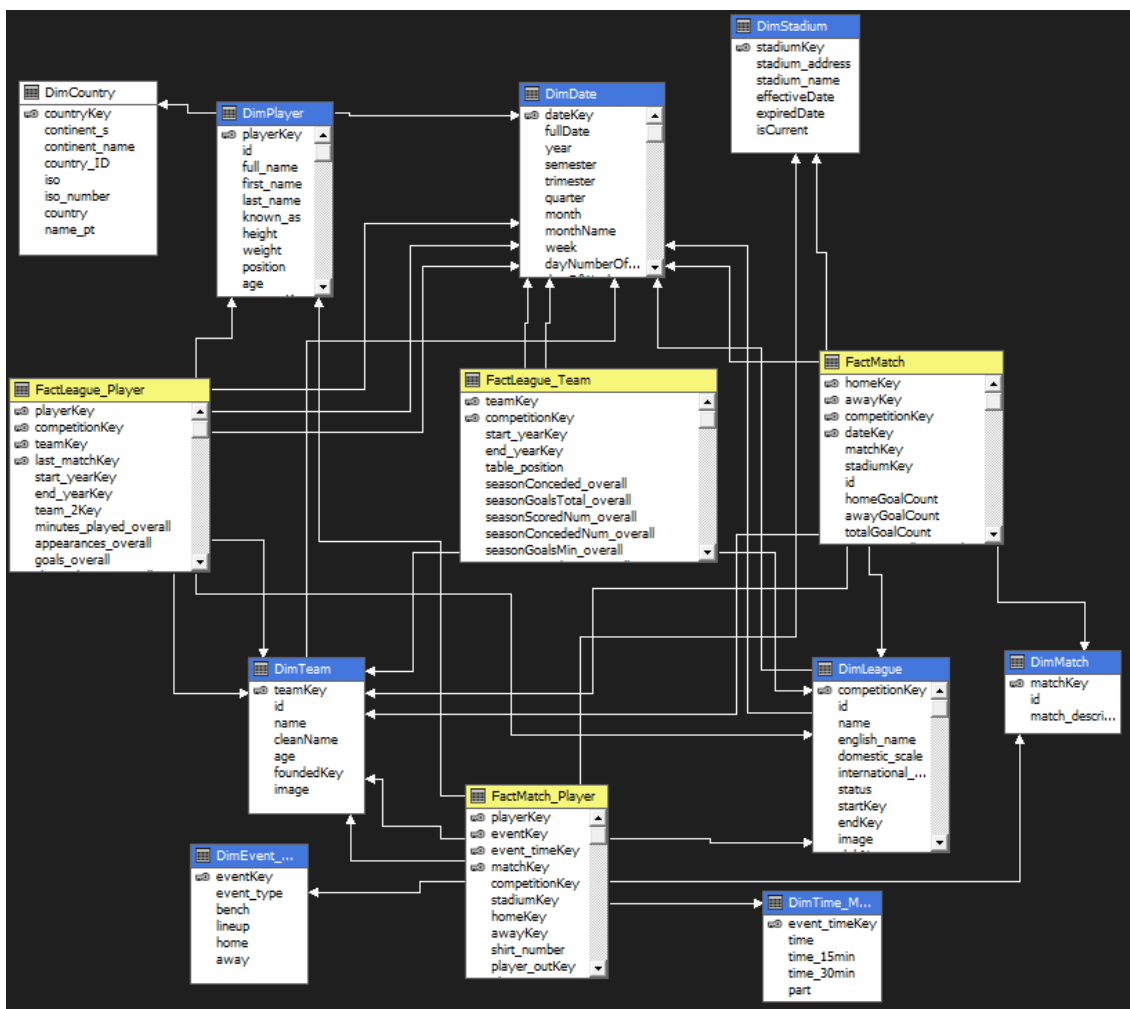


Figura 64 – Cubo OLAP

Com o propósito de simplificar a avaliação de uma disposição temporal, foram definidas cinco hierarquias para os dados contidos na dimensão DimDate e uma hierarquia para a DimTime_Match. Estas hierarquias são apresentadas na Figuras 65 e 66.

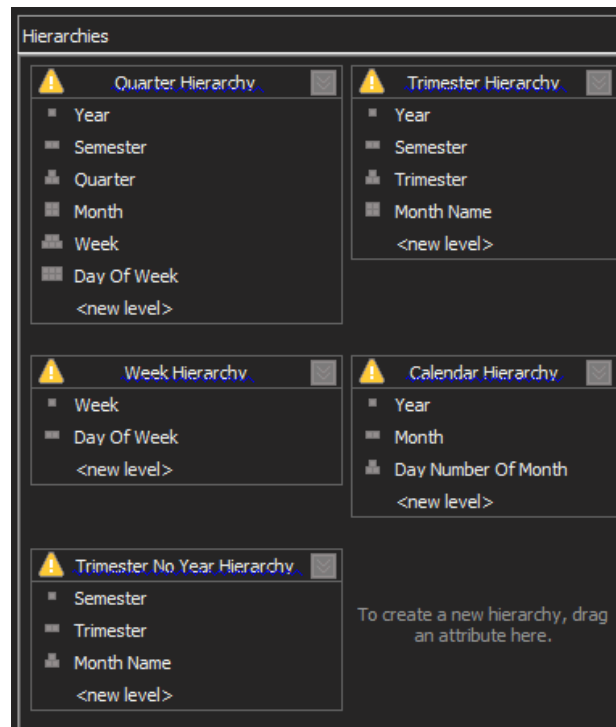


Figura 65 – Hierarquias DimDate



Figura 66 – Hierarquia DimMatch_Time

5 Análise de Resultados e Avaliação da Solução

Este capítulo está estruturado em duas partes bem definidas. Na primeira parte, são detalhados os *dashboards* desenvolvidos e as análises alcançadas. Estes *dashboards* foram cuidadosamente implementados, seguindo as diretrizes previamente estabelecidas, de forma a fornecer toda a informação relevante aos profissionais de futebol. A segunda parte descreve a avaliação da solução de *Business Intelligence* desenvolvida, incluindo a metodologia de avaliação utilizada e a análise de resultados obtidos.

5.1 Análise de Resultados

Os indicadores mencionados anteriormente na Subsecção 3.5, que são objeto de estudo ao longo desta dissertação, são apresentados através de diversos *dashboards*. Nestes, destacam-se indicadores relativos ao desempenho dos atletas nos jogos e competições, ao desempenho das equipas nas competições em que participam, bem como às prestações registadas de ambas as equipas durante uma partida de futebol. Estes indicadores permitem analisar a performance de um determinado clube e, ou jogadores, proporcionando vantagens competitivas na preparação de futuras partidas.

As análises realizadas destinam-se aos profissionais de futebol, pelo que devem ser visualmente acessíveis e de fácil interpretação, permitindo a extração rápida e eficiente de conclusões. Seguidamente são apresentados os diversos *dashboards* criados.

5.1.1 Dashboards

Os *dashboards* constituem uma ferramenta essencial para a visualização de dados, promovendo a melhoria na tomada de decisões ao amplificar a cognição e as capacidades de perceção humana. Estes oferecem uma solução eficaz para a gestão da elevada quantidade de informação disponível, introduzindo conceitos e aplicações como mapas estratégicos, *scorecards* e *Business Intelligence* (BI) [12]. Para potenciar as capacidades de visualização dos dados, foram desenvolvidos diversos *dashboards*, cada um agregando diferentes conjuntos de informação mencionados na Secção 3.5. Através destes, é possível visualizar informação desportiva, permitindo ao utilizador filtrar a informação por clube, atleta e competição.

5.1.1.1 Desempenho das Equipas

O primeiro *dashboard* apresentado na Figura 67 disponibiliza toda a informação necessária sobre o desempenho de uma equipa ao longo das várias épocas na sua respetiva competição nacional, permitindo a filtragem de dados por ano, equipa e competição.

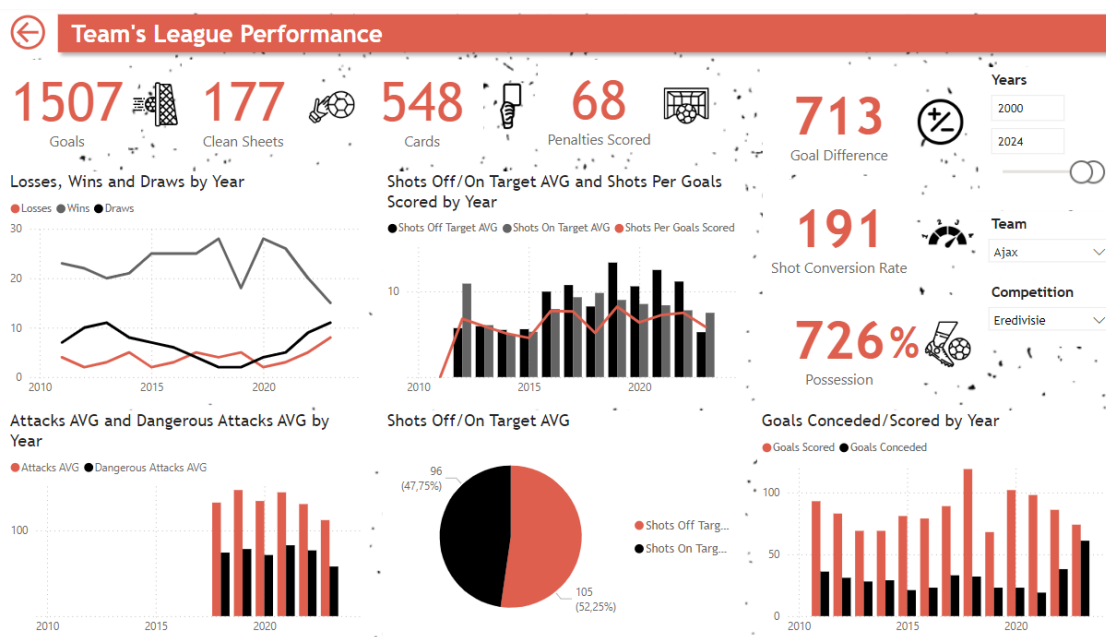


Figura 67 – Desempenho das Equipas

Conforme apresentado, verifica-se que o clube Ajax participou na principal competição dos Países Baixos (Eredivisie) ao longo de todo o século XXI, tendo marcado 1507 golos, dos quais 68 resultaram de grandes penalidades. Adicionalmente, o clube acumulou 548 cartões e completou 177 jogos sem sofrer golos, registando uma diferença positiva de 713 golos.

Ao longo do tempo, tem-se observado um aumento no número de vitórias e uma diminuição nos empates e derrotas, com exceção dos últimos anos, em que se verifica uma inversão dessa tendência, indicando um período de menor desempenho da equipa de Amesterdão. Em relação ao nível médio de ataques e ataques perigosos, estes têm variado, apresentando uma clara diminuição nos anos recentes. No que se refere ao nível médio de remates desenquadrados e enquadrados, observa-se uma proximidade entre ambos. Contudo, ao analisar os valores numa perspetiva temporal, verifica-se um aumento significativo desde o ano de 2015, seguido de uma diminuição acentuada desses indicadores em épocas mais recentes. Esta tendência culmina no ano de 2024, onde se regista um maior número de remates enquadrados em comparação com os desenquadrados, resultando num número de remates por golo praticamente idêntico ao do ano de 2018, mas com menos remates. Por fim, observa-se uma nítida diminuição no número de golos marcados e um aumento no número de golos sofridos, culminando num valor histórico máximo de 61 golos sofridos na presente época.

5.1.1.2 Desempenho dos Atletas

O segundo *dashboard*, apresentado na Figura 68, contém toda a informação crucial para avaliar o desempenho de um atleta ao longo das diversas épocas que atuou, podendo filtrar-se a informação por ano, jogador e competição.



Figura 68 – Desempenho dos Atletas

Conforme o apresenta, Aaron Lennon, atualmente com 37 anos de idade, 1,65 metros de altura e um peso de 63 quilogramas, retirado dos relvados profissionais, iniciou parte significativa da sua formação no Tottenham Hotspur FC, um clube de Londres, onde fez o maior número de aparições, assistências, golos e recebeu o maior número de cartões. Posteriormente, teve passagens pelo Everton FC e Burnley FC, ambos clubes também da Premier League, e uma experiência internacional no Kayserispor, atuando na liga turca (Süper Lig).

Durante a sua carreira profissional, não marcou nenhum golo de grande penalidade, recebeu 1 cartão vermelho e 25 cartões amarelos, acumulando um total de 27.670 minutos em campo. Em 112 partidas, a sua equipa manteve a baliza inviolada. O seu desempenho variou ao longo do tempo em termos de aparições, golos e assistências, atingindo o auge de assistências (10) em 2009 e de golos (5) em 2015. Em média, Aaron esteve envolvido em mais de 3 golos por cada 90 minutos jogados e marcou pouco mais de 1 golo por 90 minutos. A variação dos golos sofridos pela equipa onde atuou foi mais pronunciada na fase final da sua carreira.

5.1.1.3 Resumo das Partidas

Por fim, o último *dashboard* agrega toda a informação indispensável para a avaliação de um confronto entre duas equipas da mesma competição nacional (Figura 69). Este permite filtrar a informação por época, equipa visitada e visitante.

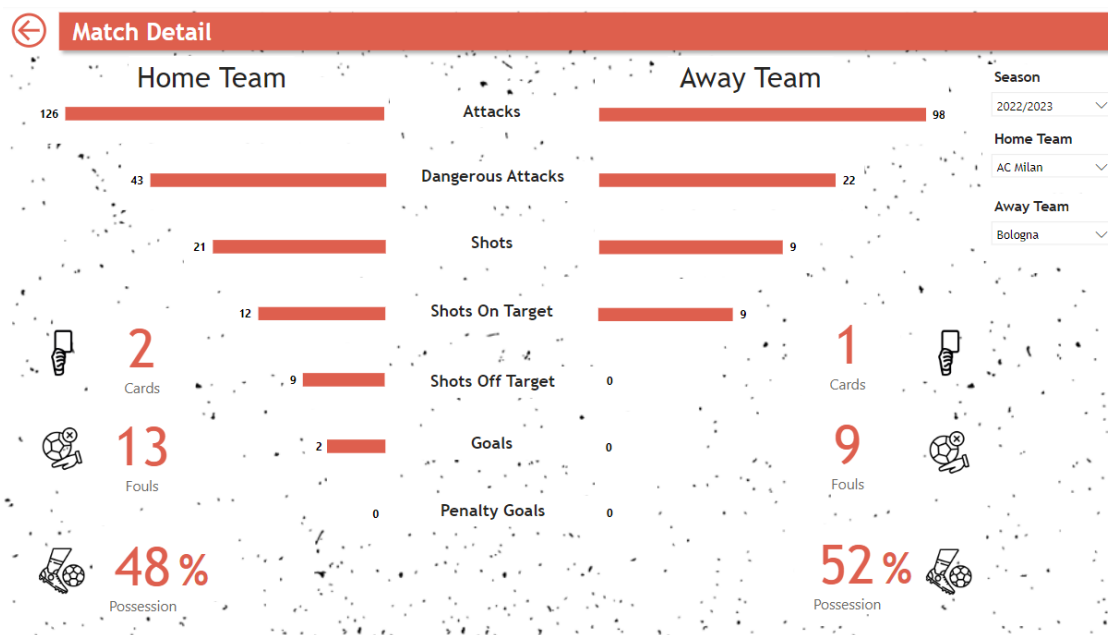


Figura 69 – Resumo das Partidas

Conforme se pode observar no confronto entre dois clubes da primeira divisão italiana, o AC Milan e o Bologna, o jogo culminou numa vitória em casa por 2-0 para a equipa de Milão, que demonstrou um domínio claro em todos os aspetos do jogo, exceto na posse de bola. O AC Milan realizou um total de 126 ataques, sendo 43 deles perigosos, com 21 remates ao alvo, dos quais 12 foram certos. Não houve golos marcados de grande penalidade e foram cometidas 13 faltas, resultando em 2 cartões. Por sua vez, o Bologna não conseguiu acompanhar o ritmo do AC Milan, somando apenas 98 ataques, dos quais apenas 22 foram considerados perigosos. O Bologna realizou 9 remates, todos no alvo. A eficácia do AC Milan foi o fator determinante para o resultado da partida.

5.1.2 Interface Mobile

De modo a assegurar uma utilização eficiente e intuitiva para os diversos utilizadores, independentemente do dispositivo utilizado, promove-se uma maior acessibilidade e flexibilidade na análise de informações através da replicação dos *dashboards* para uma versão móvel. A Figura 70 apresenta um excerto dos três *dashboards* supramencionados na sua versão móvel.

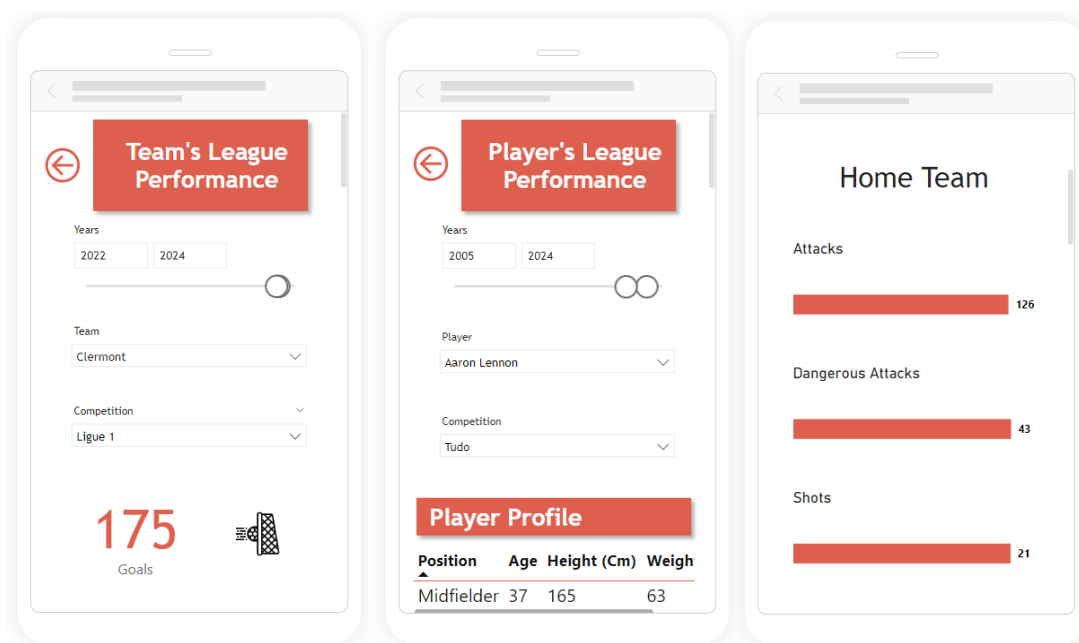


Figura 70 – Interface mobile

5.2 Avaliação da Solução

Esta secção tem como objetivo avaliar o sistema de *Business Intelligence* dedicado à análise de dados de futebol. Para tal, utiliza-se um conjunto de grandezas para aferir a viabilidade da solução implementada. A avaliação da solução foi conduzida através de um inquérito de satisfação, nos quais são analisados requisitos não funcionais como a usabilidade, integridade dos dados, desempenho, compatibilidade e histórico. A modularidade, um requisito que não pode ser avaliado por questionário, é validado através da arquitetura apresentada na Secção 3.8.

5.2.1 Metodologia de Avaliação

A solução criada foi avaliada por meio de um questionário de satisfação aplicado aos membros da equipe técnica do RANS Nusantara envolvidos diretamente no seu desenvolvimento. O propósito deste inquérito era obter a opinião dos futuros utilizadores regulares da ferramenta, a fim de entender como ela pode aprimorar o processo de preparação de futuros jogos e identificar possíveis áreas de melhoria. Os profissionais de futebol possuem habilidades analíticas afiadas para avaliar a importância de relatórios e métricas na procura por vantagens competitivas, considerando as circunstâncias que enfrentam diariamente.

O questionário, que se encontra no Anexo 6, foi disponibilizado e conduzido de forma *online*, o que representa uma maneira ágil e intuitiva de recolher as respostas dos profissionais. Este método também possibilita usar as ferramentas disponíveis na Internet para obter rapidamente as avaliações das respostas recolhidas. No inquérito de satisfação, foram utilizadas diversas afirmações categorizadas de acordo com a escala de Likert, que avalia a satisfação do utilizador em vários níveis. O método adotado utilizou cinco níveis de satisfação para analisar como os utilizadores se comportam em relação à concordância com as declarações, à avaliação do serviço e à frequência de uso. A Tabela 9 contém a escala usada nos questionários de satisfação.

Tabela 9 – Escala de Likert

Escala	Descrição
1	Discordo Totalmente
2	Discordo
3	Neutro
4	Concordo
5	Concordo Totalmente

5.2.2 Resultados dos Inquéritos

De acordo com o anteriormente mencionado, foi realizado um inquérito de satisfação à equipa técnica de profissionais de futebol da RANS Nusantara, contando com a participação de 7 elementos. Este inquérito consistiu inicialmente num conjunto de questões destinadas ao levantamento de dados estatísticos, seguido de um segundo grupo de questões para validar os requisitos funcionais e não funcionais da solução desenvolvida. De acordo com as opiniões recolhidas, os resultados obtidos foram, em geral, bastante positivos, conforme ilustrado graficamente no Anexo 7. A Tabela 10 sintetiza a frequência das respostas do segundo grupo para cada questão.

Tabela 10 – Frequência de respostas

Questão/Resultado	1	2	3	4	5
Q1	0% (0)	0% (0)	0% (0)	71,4% (5)	28,6% (2)
Q2	0% (0)	0% (0)	0% (0)	42,9% (3)	57,1% (4)
Q3	0% (0)	0% (0)	14,3% (1)	14,3% (1)	71,4% (5)
Q4	0% (0)	0% (0)	0% (0)	71,4% (5)	28,6% (2)
Q5	0% (0)	0% (0)	0% (0)	57,1% (4)	42,9% (3)
Q6	0% (0)	0% (0)	28,6% (2)	14,3% (1)	57,1% (4)
Q7	0% (0)	0% (0)	0% (0)	57,1% (4)	42,9% (3)
Q8	0% (0)	0% (0)	0% (0)	42,9% (3)	57,1% (4)
Q9	0% (0)	0% (0)	14,3% (1)	14,3% (1)	71,4% (5)
Q10	0% (0)	0% (0)	0% (0)	42,9% (3)	57,1% (4)
Q11	0% (0)	0% (0)	14,3% (1)	57,1% (4)	28,6% (2)
Q12	0% (0)	0% (0)	0% (0)	71,4% (5)	28,6% (2)

Conforme evidenciado pelos resultados, na primeira questão, cinco participantes selecionaram o nível de concordância 4 (Concordo) e dois selecionaram 5 (Concordo Totalmente). Conclui-se assim que os dashboards desenvolvidos são intuitivos. Na segunda questão, quatro participantes responderam 4 (Concordo), refletindo a opinião da maioria, enquanto os restantes três participantes responderam 5 (Concordo Totalmente). Também neste ponto verifica-se que a maioria concorda que a solução possui uma transição suave entre os diversos dashboards desenvolvidos.

A terceira questão contou com um participante a responder 3 (Neutro), a mesma quantidade a responder 4 (Concordo), enquanto a maioria, ou seja, os restantes cinco participantes, responderam 5 (Concordo Totalmente). Assim, confirmando-se que os indicadores apresentados nos dashboards são facilmente compreensíveis. Na quarta questão, os participantes mostraram concordância, com cinco a responder 4 (Concordo) e dois a responder 5 (Concordo Totalmente). Deste modo, a maioria está de acordo com a acuidade e confiabilidade das informações apresentadas.

A quinta questão foi respondida por quatro participantes com 4 (Concordo) e três com 5 (Concordo Totalmente), mostrando concordância com a periodicidade de atualização das informações exibidas. Na sexta questão, dois participantes responderam 3 (Neutro), um respondeu 4 (Concordo) e os restantes quatro responderam 5 (Concordo Totalmente), indicando que a maioria está de acordo com a integridade das informações apresentadas.

Na sétima questão, quatro participantes responderam 4 (Concordo) e três responderam 5 (Concordo Totalmente), indicando que o tempo de resposta dos dashboards é bastante satisfatório. Na oitava questão, três participantes responderam 4 (Concordo), enquanto a maioria, ou seja, os restantes quatro participantes, respondeu 5 (Concordo Totalmente). Tal como na questão anterior, verifica-se que não existem atrasos na interação com os dashboards.

A nona questão teve um participante a responder 3 (Neutro), a mesma quantidade a responder 4 (Concordo), enquanto a maioria, ou seja, os restantes cinco participantes, respondeu 5 (Concordo Totalmente). Confirma-se assim que os indicadores foram apresentados corretamente em diversos dispositivos. Na décima questão, três participantes responderam 4 (Concordo), enquanto a maioria, ou seja, os restantes quatro participantes, respondeu 5 (Concordo Totalmente). Semelhante à questão anterior, verifica-se que os indicadores foram apresentados corretamente nos diversos sistemas operativos dos dispositivos utilizados.

A décima primeira questão teve um participante a responder 3 (Neutro), enquanto a maioria, ou seja, quatro participantes, respondeu 4 (Concordo), e os restantes dois responderam 5 (Concordo Totalmente). Posto isto, afirma-se que os dados históricos são suficientes para análise, embora possam ser melhorados. Por último, na décima segunda questão, cinco participantes responderam 4 (Concordo), refletindo a maioria, enquanto dois participantes responderam 5 (Concordo Totalmente). A última questão evidencia que o histórico de dados é bem representativo das tendências de desempenho, tanto das equipas quanto dos atletas.

5.2.3 Conclusão

A análise dos resultados revela uma concordância geral entre os participantes de que os dashboards são intuitivos (Questão 1), a navegação é fácil (Questão 2) e os indicadores incluídos são compreensíveis (Questão 3), validando assim o requisito de **Usabilidade** estabelecido na Subsecção 3.4.2.2.

Além disso, a Questão 4, que confirma a precisão e confiabilidade dos dados, assegura o cumprimento do requisito de **Integridade dos Dados**. A concordância sobre a frequência necessária de atualização dos dados (Questão 5), a integridade dos dados históricos (Questão 6), a suficiência das informações históricas para análises (Questão 11) e a identificação de tendências (Questão 12) valida o requisito de **Histórico**. A maioria dos respondentes concorda

com o tempo de resposta ao carregar a informação (Questão 7) e a ausência de atrasos na disponibilização de informações (Questão 8), o que valida o requisito de **Desempenho**.

Por fim, as Questões 9 e 10, nas quais os participantes concordaram com a boa adaptação dos dashboards aos dispositivos e sistemas operativos utilizados, confirmam o requisito de **Compatibilidade**.

Em suma, ainda que os inquéritos de satisfação tenham sido respondidos por uma amostra limitada de profissionais do futebol, os resultados obtidos evidenciam a utilidade da solução de *Business Intelligence* na execução das tarefas diárias. Contudo, foram registadas algumas respostas que não concordavam com as questões apresentadas, o que indica que há margem para melhorias na ferramenta. Estes *feedbacks* são cruciais para orientar futuras melhorias e refinamentos, garantindo uma evolução contínua e a adaptação da solução às necessidades dos profissionais de futebol.

6 Conclusão

Neste capítulo são apresentadas as conclusões da presente dissertação. São descritos os objetivos alcançados, bem como os problemas e limitações descobertos ao longo da sua realização. Por fim, são identificadas potenciais oportunidades a serem exploradas em futuros trabalhos.

6.1 Objetivos Alcançados

Esta dissertação teve como objetivo o desenvolvimento de um sistema de *Business Intelligence* (BI) que proporcionasse vantagens competitivas no futebol, fornecendo toda a informação necessária aos profissionais da área, de modo a apoiar o processo de tomada de decisão. No decurso deste trabalho, foi contextualizado o problema, realizado um estudo sobre o futebol, analisado o contributo dos sistemas de informação para esta área, abordados conceitos e explorada a aplicabilidade de sistemas de BI, considerando o tema em estudo.

Adicionalmente, foram analisadas várias ferramentas de ETL e BI para identificar as mais adequadas às necessidades. As ferramentas selecionadas foram o SQL Server Integration Services (SSIS) e o Power BI da Microsoft. A combinação destas ferramentas proporciona melhor performance, facilidade de integração e interatividade para o utilizador. O sistema de BI possui uma arquitetura centralizada, preparada para situações futuras, garantindo a disponibilização de informação segundo um único padrão e capaz de armazenar histórico. Para assegurar a melhor qualidade de informação para os profissionais de futebol, foi também garantida a possibilidade de validar e filtrar informação, bem como a possibilidade de configurar

acessos diferenciados para aceder aos *dashboards* desenvolvidos relativos a clubes, atletas e competições, garantindo-se assim o cumprimento dos requisitos funcionais delineados. Este estudo demonstra que soluções deste tipo podem contribuir significativamente para futuras inovações na área.

Finalmente, para validar o cumprimento dos requisitos não funcionais mencionados nesta dissertação, foram realizados inquéritos de satisfação à equipa técnica da época 2023/2024 do Rans Nusantara. Os formulários disponibilizados visaram obter opiniões sobre a usabilidade do sistema, o grau de satisfação relativo ao tempo de resposta, o cumprimento dos requisitos e a importância dos indicadores implementados. O inquérito de satisfação realizado junto dos profissionais de futebol revelou opiniões bastante positivas, indicando que os objetivos definidos para a aplicação foram alcançados.

6.2 Problemas/Limitações

Apesar de os objetivos enunciados na presente dissertação terem sido atingidos, diversos problemas e limitações surgiram ao longo da sua realização.

Primeiramente, a API da Footystats apresentou várias limitações. A primeira foi a restrição de 1800 chamadas por hora aos diversos *endpoints*, mesmo com a subscrição ativa. Esta limitação dificultou a extração de dados, levando à necessidade de criar uma tabela de cache relativa à tabela Jogador_Jogo para contornar o problema. Para preencher essa tabela, era necessário invocar o *endpoint* individualmente para cada identificador de jogo, e, devido a essa limitação, tornou-se impossível realizar a extração continuamente sem recorrer a uma recolha prévia dos dados em cache.

Em segundo lugar, a API não disponibilizava *endpoints* para a extração de informações exclusivas de jogadores e equipas, fornecendo apenas dados no contexto das competições. Isso resultou na necessidade de remover duplicados durante a extração desses dados.

Além disso, a API apresentava diversas inconsistências nos dados. Por exemplo, não disponibilizava informações sobre os treinadores dos clubes durante as competições, embora incluísse os identificadores dos mesmos nas tabelas de Jogos e Clubes. Adicionalmente, das competições selecionadas, algumas forneciam informações desde a época 2010/2011, enquanto outras apenas disponibilizavam dados a partir da época 2015/2016.

No âmbito dos *dashboards* desenvolvidos, não foi possível traduzir os endereços em imagens, o que levou à sua exclusão dos mesmos.

Por fim, identificou-se um problema menor relacionado com a API, que fornecia arrays aninhados, dificultando a transposição dos dados extraídos para uma tabela.

6.3 Trabalho Futuro

Seguem-se algumas melhorias possíveis para futuras funcionalidades do projeto, para além das indicadas pelos inquiridos aquando dos questionários de satisfação.

Criação de mais *Dashboards*: Embora a análise de dados tenha sido extensa, o potencial da fonte de dados está longe de ser completamente explorado. Cada tabela de factos inicial conta com mais de 100 atributos, dos quais, devido a restrições de tempo, utilizou-se menos de metade, selecionando apenas os mais valiosos. No entanto, os atributos não utilizados podem complementar os atuais, proporcionando mais contexto ao utilizador final.

Inclusão de Tabelas de Facto: Em consonância com o ponto anterior, foram selecionadas apenas as tabelas de factos mais relevantes para os utilizadores finais, resultando na exclusão da tabela de factos "Árbitros_Competições".

Inclusão de mais competições: Atualmente, a ferramenta abrange exclusivamente as dez melhores competições, de acordo com o ranking da UEFA, distribuída por diversas temporadas. A inclusão de mais competições seria benéfica para ampliar a diversidade dos dados, oferecendo uma visão mais abrangente e global do futebol. Contudo, a limitação de tempo necessário para a aquisição dos dados permanece um desafio.

Tratamento dos DQPs: Dada a existência de variados DQPs, um ponto de melhoria seria analisá-los e tratá-los de forma a incluir maior volume de dados para as análises.

Integração com outras fontes de dados: Além dos dados relativos às competições, atletas, jogos e equipas, é possível integrar dados provenientes de outras fontes, como dados de GPS sobre o posicionamento dos atletas durante as partidas, dados coletados durante os treinos, indicadores biométricos (como horas de sono e níveis de cansaço) e informações sobre nutrição. Esta abordagem proporcionaria um contexto mais abrangente sobre o mundo do futebol, permitindo uma melhor preparação para obter vantagens competitivas.

Referências

- [1] S. S. Ali Fathima J and S. V P, “Data analytics in football sport to identify gaps for the improvement of quality opportunities throughout world-wide teams,” *International Journal of Recent Technology and Engineering*, pp. 2277–3878, 2018.
- [2] M. S. Domingues, “Uso de business intelligence como ferramenta de apoio em previsões de apostas de jogos de futebol,” Brasil, 2021.
- [3] D. Montes García and others, “El uso de las tecnologías BD para la gestión económica del fútbol,” 2023.
- [4] B. B. D. C. Pinto, “How can the usage of data analytics help football clubs create a competitive advantage: a research on Liverpool football club,” 2021.
- [5] E. Giorgio, “Data analytics e aziende calcistiche: gli impatti sulla gestione sportiva ed economica: i casi di successo del Brentford, Midtjylland e Liverpool,” 2021.
- [6] E. M. Rocha-Lima, “O uso do jogo eletrônico Football Manager e do portal Transfermarkt para mensurar o potencial de futebolistas promissores,” 2022.
- [7] IFAB, “Laws of the Game 23/24,” in *IFAB*, 2023.
- [8] “Tournaments.” Accessed: Oct. 29, 2023. [Online]. Available: <https://www.fifa.com/fifaplus/en/tournaments>
- [9] L. M. S. Fontanive, “Football analytics: a literature analysis from 2010 to 2020,” 2021.
- [10] D. M. C. P. Cabral, “The influence of data analysis on football teams to increase sports’ performance,” 2023.
- [11] A. de Miguel Toquero and others, “Implementación de una herramienta de análisis en la gestión deportiva mediante power bi,” 2020.
- [12] S. C. A. Ferreira, “Sistema distribuído para recolha de dados em eventos desportivos,” 2020.
- [13] R. Bhatnagar and M. Babbar, “A systematic review of sports analytics,” *International Journal of Technology Transfer and Commercialisation*, vol. 19, no. 4, pp. 393–406, 2022.
- [14] S. Diel, S. Ifland, F. Wytopil, and C. Buck, “How Digital Technologies Transform Football-A Structured Literature Review.,” *PACIS*, p. 81, 2021.
- [15] F. Ahmed Shah, M. Kretzer, and A. Mädche, “Designing an Analytics Platform for Professional Sports Teams,” 2015.

- [16] T. Blobel and M. Lames, "A Concept for Club Information Systems (CIS) An Example for Applied Sports Informatics," *Int J Comput Sci Sport*, vol. 19, no. 1, pp. 102–122, Jul. 2020, doi: 10.2478/ijcss-2020-0006.
- [17] S. A. G. Cunha, "Data Analytics in the Football Industry: The Portuguese Football Federation Case," 2022.
- [18] S. V. Baeza Carrasco, "Sistema BI de datos de fútbol," 2023.
- [19] J. vom Brocke, A. Hevner, and A. Maedche, "Introduction to Design Science Research," 2020, pp. 1–13. doi: 10.1007/978-3-030-46781-4_1.
- [20] R. J. da R. Lima, "Extração e análise multidimensional de dados de atletismo a partir de dados não estruturados," 2018.
- [21] T. A. P. Filipe, "Data Analytics Mechanisms for Sport's Related Data Retrieval and Multimedia Augmentation," 2019.
- [22] A. de C. D. Ferreira, "Um modelo para a gestão de informações do esporte de alto rendimento no Brasil," 2018.
- [23] A. L. González Ardevínez, "Cuadro de mando para el análisis de 'laliga,'" 2022.
- [24] E. Vermeulen and S. Venkata Yadavalli, "Big data in sport analytics: applications and risks," 2018.
- [25] C. N. Fontanillas, Í. Galdino De Souza, M. De, S. Leão, H. J. Santos, and R. Cosenza, "BUSINESS ANALYTICS: AN EXPLANATORY REPORT ON THE EVOLUTION OF BUSINESS INTELLIGENT AS AN ADVANTAGE IN THE DECISION-MAKING PROCESS," *Sustainable Business International Journal*, vol. 1, no. 95, 2023.
- [26] P. Cortez and A. Bifet, "Fifth special issue on knowledge discovery and business intelligence," 2020.
- [27] V. Ratten and G. Dickson, "Big data and business intelligence in sport," 2020, pp. 25–35. doi: 10.4324/9780367854454-3.
- [28] F. S. R. Santos and J. Schimiguel, "Análise de desempenho e perfis de times do campeonato brasileiro," *Cuadernos de Educación y Desarrollo*, no. 105, 2019.
- [29] R. H. V. S. Alcarva, "Estágio Curricular nos Departamentos de Data Science e Software Development no Sport Lisboa e Benfica," Instituto Superior de Economia e Gestão, 2021.
- [30] A. J. Palm, "Using business intelligence to analyse sport association's financial data," University of Twente, 2021.

- [31] M. L. Álvarez Ibarra and K. R. Vélez Cusme, “Implementación de un sistema de inteligencia de negocios en el departamento de ventas del Almacén Deportivo CTL SPORTS mediante la aplicación de herramientas BI para un adecuado proceso de toma de decisiones gerenciales.,” Ecuador: La Mana: Universidad Técnica de Cotopaxi (UTC), 2022.
- [32] “Tabelas de fatos - Documentação da IBM.” Accessed: Nov. 09, 2023. [Online]. Available: <https://www.ibm.com/docs/pt-br/tap/3.5.1?topic=structures-fact-tables>
- [33] “Conceitos de Tabelas de Dimensão e Fato na Construção do Data Warehouse | LinkedIn.” Accessed: Nov. 12, 2023. [Online]. Available: <https://www.linkedin.com/pulse/conceitos-de-tabelas-dimens%C3%A3o-e-fato-na-constru%C3%A7%C3%A3o-do-carlos/?originalSubdomain=pt>
- [34] A. Qaiser, M. U. Farooq, S. Muhammad, N. Mustafa, and N. Abrar, “Comparative Analysis of ETL Tools in Big Data Analytics,” *Pakistan Journal of Engineering and Technology*, vol. 6, no. 1, pp. 7–12, 2023.
- [35] “Magic Quadrant Research Methodology | Gartner.” Accessed: Nov. 13, 2023. [Online]. Available: <https://www.gartner.com/en/research/methodologies/magic-quadrants-research>
- [36] “Gartner Reprint 2.” Accessed: Nov. 13, 2023. [Online]. Available: https://www.gartner.com/doc/reprints?id=1-2AWA2A6O&ct=220822&st=sb&mkt_tok=MDQ5LURLSy03OTYAAAGPZUKAXfCLAmyOk_wk3xTi8IXUiBYXviG7GJA2LDioz9C7W9Cp3C1FNSfrgIChMDAs2HoEmirU-9hg9uGVMNeCzKVxS7BBoAeYip3v3zARwno
- [37] “2022 Gartner® Magic Quadrant™ for Data Integration Tools | Qlik®.” Accessed: Nov. 12, 2023. [Online]. Available: <https://www.qlik.com/us/gartner-magic-quadrant-for-data-integration-tools?ga-link=gmqdi23-comingsoon-hero>
- [38] “What is Talend ? — An Introduction to Talend ETL Tool | by Swatee Chand | Edureka | Medium.” Accessed: Nov. 14, 2023. [Online]. Available: <https://medium.com/edureka/what-is-talend-tool-881b41fc1f5f>
- [39] “(3) Is it possible use git with Talend Open Studio? | LinkedIn.” Accessed: Nov. 14, 2023. [Online]. Available: <https://www.linkedin.com/pulse/possible-use-git-talend-open-studio-maurice-dongmo/>
- [40] “Pentaho Data Integration - Hitachi Vantara Lumada and Pentaho Documentation.” Accessed: Nov. 14, 2023. [Online]. Available: https://help.hitachivantara.com/Documentation/Pentaho/Data_Integration_and_Analytics/9.1/Products/Pentaho_Data_Integration

- [41] “Capítulo 8 SSIS (SQL Server Integration Services) | Sistemas Multidimensionales.” Accessed: Nov. 14, 2023. [Online]. Available: <https://jsamos.es/sm2019/etl-ssis.html>
- [42] “Gartner Reprint.” Accessed: Nov. 12, 2023. [Online]. Available: https://www.gartner.com/doc/reprints?id=1-2955ETOT&ct=220215&st=sb?ocid=lp_pg398450_gdc_comm_az
- [43] “Introdução aos painéis para designers do Power BI - Power BI | Microsoft Learn.” Accessed: Nov. 14, 2023. [Online]. Available: <https://learn.microsoft.com/pt-pt/power-bi/create-reports/service-dashboards>
- [44] T. Gunadham, “A Study of Business Intelligence Tools from Users’ Perspectives,” *Interdisciplinary Research Review*, vol. 18, no. 3, pp. 8–14, 2023.
- [45] “Qlik Sense Teste Gratuito | Qlik.” Accessed: Nov. 14, 2023. [Online]. Available: <https://www.qlik.com/pt-br/trial/qlik-sense-business>
- [46] “Tableau 2022.1 new features.” Accessed: Nov. 14, 2023. [Online]. Available: <https://www.tableau.com/en-gb/2022-1-features>
- [47] R. Sharda, D. A. Asamoah, and N. Ponna, “Research and pedagogy in business analytics: Opportunities and illustrative examples,” *Journal of Computing and Information Technology*, vol. 21, no. 3, pp. 171–183, 2013, doi: 10.2498/cit.1002194.
- [48] T. advisor Tomi Seppälä, “Utilization of Business Analytics in Professional Sports Organizations Title of thesis Utilization of Business Analytics in Professional Sports Organizations,” 2020. [Online]. Available: www.aalto.fi
- [49] D. Schrader, A. Gupta, L. S. Iyer, S. Schiller, and R. Sharda, “Sports Analytics Research Collaborations: Connecting Business Schools with Athletic Departments.,” in *AMCIS*, 2016.
- [50] O. Caya and A. Bourdon, “A framework of value creation from business intelligence and analytics in competitive sports,” in *Proceedings of the Annual Hawaii International Conference on System Sciences*, IEEE Computer Society, Mar. 2016, pp. 1061–1071. doi: 10.1109/HICSS.2016.136.
- [51] N. M. Watanabe, S. Shapiro, and J. Drayer, “Big data and analytics in sport management,” *Journal of Sport Management*, vol. 35, no. 3, pp. 197–202, May 2021, doi: 10.1123/JSM.2021-0067.
- [52] J. H. M. M. Guimarães, “Data analytics applied to football and football players,” 2018.
- [53] A. S. Rouiller, “Plataforma de agregação e análise de informação institucional e nas redes sociais: uma aplicação às camadas jovens de futebol,” 2020.
- [54] “Top Sports Analytics Companies 2023.” Accessed: Nov. 02, 2023. [Online]. Available: <https://brandessenceresearch.com/blog/sports-analytics-companies>

- [55] “Catapulta.” Accessed: Nov. 26, 2023. [Online]. Available: <https://www.catapult.com/pt/>
- [56] J. Reece and S. Y. Hong, “Big data analytics for smart sports using apache spark,” *Issues in Information Systems*, vol. 22, no. 3, pp. 1–13, 2021, doi: 10.48009/3_iis_2021_1-15.
- [57] F. A. T. Mota, “Recolha de Informação de Eventos Desportivos Usando Técnicas de Mineração de Dados: Investigação, Desenho e Implementação de uma Solução web Baseada na Nuvem,” Universidade da Beira Interior (Portugal), 2020.
- [58] “Top 8 best Football stats and analysis websites.” Accessed: Nov. 26, 2023. [Online]. Available: <https://www.thepunterspage.com/useful-sites-football-betting/>
- [59] A. Battassini, “Football Stats: uma plataforma web para visualização de estatísticas e previsões sobre campeonatos de futebol,” 2022.
- [60] “Mercado de transferências, rumores, Valores de mercado e notícias | Transfermarkt.” Accessed: Nov. 27, 2023. [Online]. Available: <https://www.transfermarkt.pt/>
- [61] “El cuadro de mandos que ayuda a los clubes de LaLiga a evolucionar | Fútbol Global.” Accessed: Nov. 30, 2023. [Online]. Available: <https://newsletter.laliga.es/futbol-global/los-paneles-de-datos-que-ayudan-a-los-clubes-de-laliga-a-evolucionar-1>
- [62] “Premier League Table, Form Guide & Season Archives.” Accessed: Dec. 16, 2023. [Online]. Available: <https://www.premierleague.com/tables>
- [63] “Brighton | Stats | UEFA Europa League 2023/24 | UEFA.com.” Accessed: Dec. 16, 2023. [Online]. Available: <https://www.uefa.com/uefaeuropaleague/clubs/2601105--brighton/statistics/>
- [64] “Jupiler Pro League Ranking | Pro League | Official Web.” Accessed: Dec. 16, 2023. [Online]. Available: <https://www.proleague.be/jpl-ranking>
- [65] “Union SG | UEFA Europa League 2023/24 | UEFA.com.” Accessed: Dec. 16, 2023. [Online]. Available: <https://www.uefa.com/uefaeuropaleague/clubs/64125--union-sg/>
- [66] “Union SG | UEFA Conference League 2023/24 | UEFA.com.” Accessed: Jun. 10, 2024. [Online]. Available: <https://www.uefa.com/uefaeuropaconferenceleague/clubs/64125--union-sg/matches/>
- [67] “Estatísticas de Futebol - FootyStats.” Accessed: Mar. 28, 2024. [Online]. Available: <https://footystats.org/pt/>
- [68] D. O. Polyane, “CENTRO DE EDUCAÇÃO SUPERIOR DE BRASÍLIA CENTRO UNIVERSITÁRIO INSTITUTO DE EDUCAÇÃO SUPERIOR DE BRASÍLIA,” 2017.
- [69] M. Luiza de Almeida Campos, “Elaboração de modelos de domínio em ontologias: a abordagem onomasiológica e a função da definição,” 2017.

- [70] "What is Unified Modeling Language (UML)?" Accessed: Mar. 24, 2024. [Online]. Available: <https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-uml/>
- [71] "Country coefficients | UEFA Coefficients | UEFA.com." Accessed: Jun. 16, 2024. [Online]. Available: <https://www.uefa.com/nationalassociations/uefarankings/country/#/yr/2024>
- [72] T. T. and Y. Y. and N. B. Bennaceur Amel and Tun, "Requirements Engineering," in *Handbook of Software Engineering*, R. N. and K. K. Cha Sungdeok and Taylor, Ed., Cham: Springer International Publishing, 2019, pp. 51–92. doi: 10.1007/978-3-030-00262-6_2.
- [73] "RANS Nusantara FC (@rans.nusantara) • fotos e vídeos do Instagram." Accessed: Apr. 23, 2024. [Online]. Available: <https://www.instagram.com/rans.nusantara/>
- [74] Dr. Abdullah and Dr. M. Huda, "EMPIRICALLY VALIDATED SIMPLICITY EVALUATION MODEL FOR OBJECT ORIENTED SOFTWARE," *International Journal of Software Engineering & Applications*, vol. 9, no. 6, Nov. 2018, doi: 10.5121/ijsea.2018.9606.
- [75] "What is FURPS+? – Business Analyst Training in Hyderabad – COEPD." Accessed: Mar. 25, 2024. [Online]. Available: <https://businessanalysttraininghyderabad.wordpress.com/2014/08/05/what-is-furps/>
- [76] "ZappySys | SSIS Components for REST API, JSON, XML, Amazon, Azure.." Accessed: Apr. 27, 2024. [Online]. Available: <https://zappysys.com/>
- [77] "SSIS PowerPack - SSIS Components & Tasks | ZappySys." Accessed: Apr. 27, 2024. [Online]. Available: <https://zappysys.com/products/ssis-powerpack/>
- [78] "SSIS Json Source Adapter (Bulk Extract,Read,Copy Json Records)." Accessed: Apr. 28, 2024. [Online]. Available: <https://zappysys.com/onlinehelp/ssis-powerpack/index.htm#page=json-source.htm>

7 Anexos

Anexo 1 – Questionário de Levantamentos de Requisitos

Este anexo exhibe o questionário de levantamento de requisitos, que serviu de base para o desenvolvimento da solução.

Questionário/ Questionnaire

B I U ↻ ✖

O panorama atual do futebol é cada vez mais marcado pela incorporação de tecnologias inovadoras que impulsionam a análise e otimização de dados relacionados com o desempenho desportivo. Neste é onde se insere o âmbito da minha dissertação, na qual estou a investigar a produção de Soluções de Business Intelligence no contexto do Futebol no decorrente curso de Mestrado de Engenharia Informática no ISEP do Porto. Certamente está a questionar-se o que são Soluções de Business Intelligence e essas são nada mais nada menos do que exemplos de Sistemas de Informação que reúnem indicadores de desempenho temporais, como golos marcados, sofridos, número de remates, entre outros. Estes dados são posteriormente tratados, organizados e apresentados em dashboards interativos permitindo que qualquer profissional ligado ao futebol possa aceder de forma clara e concisa a um conjunto de dados relevantes de forma auxiliar a tomada de decisões estratégico-táticas mais bem fundamentadas.

Para melhor compreender o domínio do futebol e, conseqüentemente, desenvolver uma solução informática que responda às necessidades dos profissionais desta área em termos de análise de dados, solicito que responda ao seguinte conjunto de questões. Estas visam focar-se especificamente em dados quantitativos puramente desportivos referentes a competições (ligas), equipas e atletas, no entanto poderá indicar que outros indicadores poderiam fazer sentido vir a ser incluídos. A sua participação é fundamental para o desenvolvimento deste projeto, no qual agradeço desde já a sua colaboração.

The current football landscape is increasingly marked by the incorporation of innovative technologies that drive the analysis and optimization of data related to sports performance. This is where the scope of my dissertation comes in, in which I am investigating the production of Business Intelligence Solutions in the context of Football in the current Master's Degree in Computer Engineering at ISEP in Porto. I'm sure you're wondering what Business Intelligence Solutions are and they are nothing more or less than examples of Information Systems that collect temporal performance indicators, such as goals scored, conceded, number of shots, among others. This data is then processed, organized and presented on interactive dashboards so that any football professional can access a clear and concise set of relevant data to help make better informed strategic and tactical decisions.

In order to better understand the field of football and, consequently, develop an IT solution that meets the needs of professionals in this area in terms of data analysis, I ask you to answer the following set of questions. These are aimed specifically at purely sporting quantitative data relating to competitions (leagues), teams and athletes, but you could indicate what other indicators might make sense to include. Your participation is fundamental to the development of this project, and I would like to thank you in advance for your co-operation.

Ex:

- (1) <https://user-images.githubusercontent.com/75815212/221917008-5c8f0380-4b44-4291-a6c3-c49cf297d8c0.png>
- (2) <https://blog.enterprisedna.co/wp-content/uploads/2023/04/Screenshot-2023-04-05-at-6.31.46-am-1024x572.png.webp>
- (3) <https://www.eng.com.br/assets/img/artigo/power-bi-dashboard-futebol-eng-dtp-multimedia.jpg>

Nome (Primeiro e Último) *
Name (First and Last)

Texto de resposta curta

Licença / License *

- Nível/Level 2 (UEFA B / Conmebol B / AFC B)
- Nível/Level 3 (UEFA A / Conmebol A / AFC A)
- Nível/Level 4 (UEFA Pro / Conmebol Pro / AFC Pro)
- Outra opção...

1. Quais seriam os profissionais de futebol que considera que mais beneficiariam com este tipo de soluções de Business Intelligence? *

1. Which football professionals do you think would benefit most from this type of business intelligence solution?

Texto de resposta longa

2. Que tipo de informações específicas um profissional de futebol desejaria visualizar nos dashboards? *

2. What specific information would a football professional want to see on dashboards?

Texto de resposta longa

3. Quais são os principais desafios enfrentados pelos profissionais de futebol ao tentar tomar decisões com base em análises de dados de desempenho dos jogadores? *

3. What are the main challenges faced by football professionals when trying to make decisions based on analyses of player performance data?

Texto de resposta longa

4. Que dados estão consecutivamente a recolher sobre os jogadores (tanto em suporte papel como informático) e como estão estruturados? *

4. What data are they consecutively collecting on players (both paper and computerised) and how is it structured?

Texto de resposta longa

5. Quais são as métricas mais importantes para avaliar o desempenho dos jogadores de futebol? (ex: passes certos, distância percorrida, eficácia de remates, etc.) *

5. What are the most important metrics for evaluating the performance of football players (e.g. correct passes, distance travelled, effectiveness of shots, etc.)?

Texto de resposta longa

6. Quais são as métricas de desempenho que podem ajudar a identificar potenciais áreas de melhoria para os atletas? *

6. Which performance metrics can help identify potential areas of improvement for players?

Texto de resposta longa

7. Quais são os fatores-chave que são considerados ao decidir o onze titular de jogadores numa partida? *

7. What are the key factors that are considered when deciding the starting eleven of players in a match?

Texto de resposta longa

...

8. Quais são os fatores-chave que influenciam o desempenho dos jogadores de futebol durante os jogos? *

8. What are the key factors that influence the performance of footballers during matches?

Texto de resposta longa

9. Quais são as métricas mais visualizadas aquando de análise de equipas e jogadores adversários? *

9. What are the most visualised metrics when analysing opposing teams and players?

Texto de resposta longa

10. Quais são as áreas específicas do jogo (por exemplo, ataque, defesa, posse de bola) que os profissionais de futebol desejam analisar com maior detalhe? *

10. What are the specific areas of the game (e.g. attack, defence, possession) that football professionals want to analyse in greater detail?

Texto de resposta longa

11. Para que propósito se viria a utilizar os dashboards? (ex: identificar áreas de melhoria para os jogadores, avaliar o desempenho da equipa em diferentes situações de jogo, etc.) *

11. For what purpose would you use the dashboards? (e.g. to identify areas of improvement for players, to evaluate the team's performance in different game situations, etc.)

Texto de resposta longa

⋮ ★

12. Qual é o nível de granularidade, isto é o nível de detalhe que gostaria de ver refletido nas informações apresentadas nos dashboards? (ex: dados individuais dos jogadores, estatísticas da equipa como um todo, análises por jogo, etc.)

12. What is the level of granularity, the level of detail you would like to see reflected in the information presented in the dashboards? (e.g. individual player data, team statistics, analyses per game, etc.)

Texto de resposta longa
.....

★

13. Quais são as suas preferências de visualização de dados? (ex: gráficos de barras, gráficos de linhas, mapas de calor, etc.)

13. What are your data visualisation preferences? (e.g. bar charts, line graphs, heat maps, etc.)

Texto de resposta longa
.....

★

14. Como podemos garantir que os dashboards desenvolvidos sejam intuitivos e fáceis de usar para os profissionais de futebol que não têm experiência técnica em análise de dados?

14. How can we ensure that the dashboards developed are intuitive and easy to use for football professionals who have no technical experience in analysing data?

Texto de resposta longa
.....

★

15. Para além dos dados mencionados no âmbito da dissertação, que outros indicadores gostaria de ver também refletidos em dashboards?

15. In addition to the data mentioned in the dissertation, what other indicators would you like to see reflected in dashboards?

Texto de resposta longa
.....

Anexo 2 – Tabelas do Modelo Relacional

Este anexo expõe as tabelas e respetivos atributos do modelo de dados relacional da API footystats.

Tabela de Continentes

Atributo	Descrição	Tipo
continent	Sigla do Continente	nvarchar(10) NOT NULL

Tabela de Países

Atributo	Descrição	Tipo
country_ID	ID do País	int
iso	Sigla do País	nvarchar (80)
iso_number	Código ISO do País	int
country	Nome do País (Inglês)	nvarchar (120) NOT NULL
name_pt	Nome do País (Português)	nvarchar (120)

Tabela de Ligas

Atributo	Descrição	Tipo
id	ID da Época/Competição/Liga	int NOT NULL
name	Nome da Liga (Português)	nvarchar(40) NOT NULL
english_name	Nome da Liga (Inglês)	nvarchar(40) NOT NULL
country	Nome do País (Inglês)	nvarchar(40) NOT NULL
domestic_scale	Ranking da Liga (País)	int NOT NULL
international_scale	Ranking da Liga (Global)	int NOT NULL
status	Estado da Época	nvarchar(50) NOT NULL
starting_year	Ano de Início da Época	int NOT NULL
ending_year	Ano de Fim da Época	int NOT NULL
continent	Continente onde a Liga é Realizada	nvarchar(10) NOT NULL
image	URL da Insígnia da Liga	nvarchar(260) NOT NULL
clubNum	Número Total de Clubes da Liga	int NOT NULL
season	Descrição	nvarchar(36) NOT NULL

Tabela de Jogos

Atributo	Descrição	Tipo
id	ID do Jogo	int NOT NULL
homeID	ID da Equipa Visitante	int NOT NULL
awayID	ID da Equipa Visitada	int NOT NULL
status	Estado do Jogo	nvarchar(40) NOT NULL
roundID	ID da Jornada	int NOT NULL
homeGoalCount	Número Total de Golos da Equipa Visitada	int NOT NULL
awayGoalCount	Número Total de Golos da Equipa Visitante	int NOT NULL

totalGoalCount	Número Total de Golos	int NOT NULL
team_a_yellow_cards	Número Total de Cartões Amarelos da Equipa Visitada	int NOT NULL
team_b_yellow_cards	Número Total de Cartões Amarelos da Equipa Visitante	int NOT NULL
team_a_red_cards	Número Total de Cartões Vermelhos da Equipa Visitada	int NOT NULL
team_b_red_cards	Número Total de Cartões Vermelhos da Equipa Visitante	int NOT NULL
team_a_shotsOnTarget	Número Total de Remates à Baliza da Equipa Visitada	int NOT NULL
team_b_shotsOnTarget	Número Total de Remates à Baliza da Equipa Visitante	int NOT NULL
team_a_shotsOffTarget	Número Total de Remates Fora do Alvo da Equipa Visitada	int NOT NULL
team_b_shotsOffTarget	Número Total de Remates Fora do Alvo da Equipa Visitante	int NOT NULL
team_a_shots	Número Total de Remates da Equipa Visitada	int NOT NULL
team_b_shots	Número Total de Remates da Equipa Visitante	int NOT NULL
team_a_fouls	Número Total de Faltas da Equipa Visitada	int NOT NULL
team_b_fouls	Número Total de Faltas da Equipa Visitante	int NOT NULL
team_a_possession	Percentagem de Posse de Bola da Equipa Visitada	int NOT NULL
team_b_possession	Percentagem de Posse de Bola da Equipa Visitante	int NOT NULL
refereeID	ID do Árbitro	int
coach_a_ID	ID do Treinador da Equipa Visitada	int
coach_b_ID	ID do Treinador da Equipa Visitante	int
stadium_name	Nome do Estádio	nvarchar(240) NOT NULL
team_a_cards_num	Número Total de Cartões da Equipa Visitada	int NOT NULL
team_b_cards_num	Número Total de Cartões da Equipa Visitante	int NOT NULL
date_unix	Data e Hora de Início do Jogo em Formato UNIX	int NOT NULL
team_a_dangerous_attacks	Número Total de Ataques Perigosos da Equipa Visitada	int NOT NULL
team_b_dangerous_attacks	Número Total de Ataques Perigosos da Equipa Visitante	int NOT NULL
team_a_attacks	Número Total de Ataques da Equipa Visitada	int NOT NULL
team_b_attacks	Número Total de Ataques da Equipa Visitante	int NOT NULL

team_a_penalties_won	Número Total de Grandes Penalidades Marcadas a Favor da Equipa Visitada	int NOT NULL
team_b_penalties_won	Número Total de Grandes Penalidades Marcadas a Favor da Equipa Visitante	int NOT NULL
team_a_penalty_goals	Número Total de Golos Marcados por Grande Penalidade da Equipa Visitada	int NOT NULL
team_b_penalty_goals	Número Total de Golos Marcados por Grande Penalidade da Equipa Visitante	int NOT NULL
team_a_penalty_missed	Número Total de Grandes Penalidades Falhadas da Equipa Visitada	int NOT NULL
team_b_penalty_missed	Número Total de Grandes Penalidades Falhadas da Equipa Visitante	int NOT NULL
competition_id	ID na Época/Competição/Liga	int NOT NULL

Tabela de Jogadores e Jogos

Atributo	Descrição	Tipo
matchID	ID do Jogo	int NOT NULL
homeID	ID da Equipa Visitante	int NOT NULL
awayID	ID da Equipa Visitada	int NOT NULL
roundID	ID da Jornada	int NOT NULL
refereeID	ID do Árbitro	int
stadium_name	Nome do Estádio	nvarchar(200)
competitionID	ID da Época/Competição/Liga	int NOT NULL
playerID	ID do Jogador	int NOT NULL
shirt_number	Número da Camisola	int NOT NULL
bench	Flag (1) Se Participou no Jogo a partir do Banco	bit
lineup	Flag (1) Se Participou no Jogo no Onze Titular	bit
home	Flag (1) Se Participou no Jogo na Equipa Visitada	bit
away	Flag (1) Se Participou no Jogo na Equipa Visitante	bit
event_type	Tipo de Evento	nvarchar(30)
event_time	Minuto em que Ocorreu o Evento	nvarchar(20)
player_out_id	ID do Jogador que foi Substituído	int
player_out_time	Minuto da Substituição	nvarchar(20)

Tabela de Jogadores e Ligas

Atributo	Descrição	Tipo
player_id	ID do Jogador	int NOT NULL
competition_id	ID da Época/Competição/Liga	int NOT NULL
starting_year	Ano de Início na Época	int NOT NULL
ending_year	Ano de Fim na Época	int NOT NULL
club_team_id	ID do Clube em que o Jogador Esteve Durante a Época	int NOT NULL
club_team_2_id	ID do Clube para qual o Jogador foi Transferido Durante a Época	int
minutes_played_overall	Número Total de Minutos Jogados do Jogador na Época	int NOT NULL
appearances_overall	Número Total de Jogos Disputados do Jogador na Época	int NOT NULL
goals_overall	Número Total de Golos Marcados do Jogador na Época	int NOT NULL
clean_sheets_overall	Número Total de Jogos sem Sofrer Golos da Equipa Deste Jogador nos Jogos que Participou na Época	int NOT NULL
conceded_overall	Número Total de Golos Sofridos Enquanto o Jogador Esteve em Campo na Época	int NOT NULL
assists_overall	Número Total de Assistências do Jogador na Época	int NOT NULL
penalty_goals	Número Total de Golos Marcados através de Grandes Penalidades do Jogador na Época	int NOT NULL
penalty_misses	Número Total de Grandes Penalidades Falhadas do Jogador na Época	int NOT NULL
penalty_success	Percentagem de Sucesso em Golos Marcados através de Grandes Penalidades do Jogador na Época	int NOT NULL
goals_involved_per_90_overall	Golos Envolvidos (Golos + Assistências) por 90 Minutos do Jogador na Época	decimal (6,3) NOT NULL
assists_per_90_overall	Assistências por 90 Minutos do Jogador na Época	decimal (6,3) NOT NULL
goals_per_90_overall	Golos Marcados por 90 Minutos do Jogador na Época	decimal (6,3) NOT NULL
clean_sheets_per_overall	Percentagem de Jogos sem Sofrer Golos Enquanto o Jogador Esteve em Campo na Época	int NOT NULL
min_per_goal_overall	Minutos por Golo Marcado do Jogador na Época	int NOT NULL
conceded_per_90_overall	Golos Sofridos por 90 Minutos do Jogador na Época	decimal (6,3) NOT NULL
min_per_conceded_overall	Minutos por Golo Sofrido do Jogador na Época	int NOT NULL

cards_overall	Número Total de Cartões do Jogador na Época	int NOT NULL
yellow_cards_overall	Número Total de Cartões Amarelos do Jogador na Época	int NOT NULL
red_cards_overall	Número Total de Cartões Vermelhos do Jogador na Época	int NOT NULL
min_per_match	Número Médio de Minutos do Jogador por Jogo na Época	int NOT NULL
min_per_card_overall	Minutos por Cartão (Amarelo / Vermelho) do Jogador na Época	int NOT NULL
min_per_assist_overall	Minutos por Assistência do Jogador na Época	int NOT NULL
cards_per_90_overall	Cartões por 90 Minutos do Jogador na Época	decimal (6,3) NOT NULL
rank_in_league_top_attackers	Posição do Jogador no Ranking de Melhores Atacantes da Época	int NOT NULL
rank_in_league_top_midfielders	Posição do Jogador no Ranking de Melhores Médios da Época	int NOT NULL
rank_in_league_top_defenders	Posição do Jogador no Ranking de Melhores Defesas da Época	int NOT NULL
rank_in_club_top_scorer	Posição do Jogador no Ranking de Melhores Marcadores da Equipa na Época	int NOT NULL
last_match_timestamp	Data e Hora de Início do Último Jogo Disputado pelo Jogador na Época em Formato UNIX	int NOT NULL

Tabela de Jogadores

Atributo	Descrição	Tipo
id	ID do Jogador	int NOT NULL
full_name	Nome Completo do Jogador	nvarchar(160) NOT NULL
first_name	Primeiro Nome do Jogador	nvarchar(80) NOT NULL
last_name	Último Nome do Jogador	nvarchar(120) NOT NULL
known_as	Nome Comum do Jogador	nvarchar(140) NOT NULL
height	Peso do Jogador (Quilogramas)	int NOT NULL
weight	Altura do Jogador (Centímetros)	int NOT NULL
position	Posição do Jogador	nvarchar(40) NOT NULL
birthday	Representação UNIX da Data de Nascimento do Jogador	int NOT NULL
nationality	Nacionalidade do Jogador	nvarchar(90) NOT NULL
continent	Continente do Jogador	nvarchar(10) NOT NULL

Tabela de Árbitros

Atributo	Descrição	Tipo
id	ID do Árbitro	int NOT NULL
full_name	Nome Completo do Jogador	nvarchar(80) NOT NULL
first_name	Primeiro Nome do Jogador	nvarchar(60) NOT NULL
last_name	Último Nome do Jogador	nvarchar(60) NOT NULL
known_as	Nome Comum do Jogador	nvarchar(80) NOT NULL
shorthand	Representação Amigável da Programação do Nome Completo do Jogador	nvarchar(80) NOT NULL
continent	Continente do Jogador	nvarchar(10) NOT NULL
birthday	Representação UNIX da Data de Nascimento	int NOT NULL
nationality	Nacionalidade do Jogador	nvarchar(30) NOT NULL

Tabela de Árbitros e Ligas

Atributo	Descrição	Tipo
referee_id	ID do Árbitro	int NOT NULL
competition_id	ID da Época/Competição/Liga	int NOT NULL
starting_year	Ano de Início na Época	int NOT NULL
ending_year	Ano de Fim na Época	int NOT NULL
appearances_overall	Número Total de Jogos Apitados pelo Árbitro na Época	int NOT NULL
wins_home	Número Total de Vitórias em Casa em Jogos Apitados pelo Árbitro na Época	int NOT NULL
wins_away	Número Total de Vitórias Fora em Jogos Apitados pelo Árbitro na Época	int NOT NULL
draws_overall	Número Total de Empates em Jogos Apitados pelo Árbitro na Época	int NOT NULL
wins_per_home	Percentagem de Vitórias em Casa em Jogos Apitados pelo Árbitro na Época	int NOT NULL
wins_per_away	Percentagem de Vitórias Fora em Jogos Apitados pelo Árbitro na Época	int NOT NULL
draws_per	Percentagem de Empates em Jogos Apitados pelo Árbitro na Época	int NOT NULL
btts_overall	Número Total de Jogos Apitados pelo Árbitro que Terminaram com BBTS na Época	int NOT NULL
btts_percentage	Percentagem de BTTS em Jogos Apitados pelo Árbitro na Época	int NOT NULL

goals_overall	Número Total de Golos Assinalados em Jogos Apitados pelo Árbitro na Época	int NOT NULL
goals_home	Número Total de Golos em Casa Assinalados em Jogos Apitados pelo Árbitro na Época	int NOT NULL
goals_away	Número Total de Golos Fora Assinalados em Jogos Apitados pelo Árbitro na Época	int NOT NULL
goals_per_match_overall	Número de Golos Assinalados por Jogos Apitados pelo Árbitro na Época	decimal (6,3) NOT NULL
goals_per_match_home	Número de Golos Assinalados em Casa por Jogos Apitados pelo Árbitro na Época	decimal (6,3) NOT NULL
goals_per_match_away	Número de Golos Assinalados Fora por Jogos Apitados pelo Árbitro na Época	decimal (6,3) NOT NULL
penalties_given_overall	Número Total de Grandes Penalidades Assinaladas em Jogos que o Árbitro Apitou na Época	int NOT NULL
penalties_given_home	Número Total de Grandes Penalidades Assinaladas em Casa em Jogos que o Árbitro Apitou na Época	int NOT NULL
penalties_given_away	Número Total de Grandes Penalidades Assinaladas Fora em Jogos que o Árbitro Apitou na Época	int NOT NULL
penalties_given_percentage_overall	Percentagem de Grandes Penalidades Assinaladas em Jogos Apitados pelo Árbitro na Época	int NOT NULL
penalties_given_percentage_home	Percentagem de Grandes Penalidades em Casa Assinaladas em Jogos Apitados pelo Árbitro na Época	int NOT NULL
penalties_given_percentage_away	Percentagem de Grandes Penalidades Fora Assinaladas em Jogos Apitados pelo Árbitro na Época	int NOT NULL
cards_overall	Número Total de Cartões Assinalados pelo Árbitro na Época	int NOT NULL
cards_home	Número Total de Cartões em Casa Assinalados pelo Árbitro na Época	int NOT NULL
cards_away	Número Total de Cartões Fora Assinalados pelo Árbitro na Época	int NOT NULL
cards_per_match_overall	Percentagem de Cartões Assinalados por Jogos Apitados pelo Árbitro na Época	decimal (6,3) NOT NULL
cards_per_match_home	Percentagem de Cartões em Casa Assinalados por Jogos Apitados pelo Árbitro na Época	decimal (6,3) NOT NULL

cards_per_match_away	Percentagem de Cartões em Casa Assinalados por Jogos Apitados pelo Árbitro na Época	decimal (6,3) NOT NULL
yellow_cards_overall	Número Total de Cartões Amarelos em Jogos Apitados pelo Árbitro na Época	int NOT NULL
red_cards_overall	Número Total de Cartões Vermelhos em Jogos Apitados pelo Árbitro na Época	int NOT NULL
min_per_goal_overall	Minutos por Golo em Jogos Apitados pelo Árbitro na Época	int NOT NULL
min_per_card_overall	Minutos por Cartão (Amarelo / Vermelho) em Jogos Apitados pelo Árbitro na Época	int NOT NULL

Tabela de Jornadas

Atributo	Descrição	Tipo
roundID	ID da Jornada	int NOT NULL
awayID	ID da Equipa Visitante	int NOT NULL
homeID	ID da Equipa Visitada	int NOT NULL
game_week	Semana de Jogo	int NOT NULL
competition_id	ID da Época/Competição/Liga	int NOT NULL

Tabela de Estádios

Atributo	Descrição	Tipo
stadium_address	Localização do Estádio	nvarchar(110)
stadium_name	Nome do Estádio	nvarchar(60)

Tabela de Equipas e Ligas

Atributo	Descrição	Tipo
team_id	ID da Equipa	int NOT NULL
table_position	Posição Atual na Tabela	int NOT NULL
competition_id	ID da Época/Competição/Liga	int NOT NULL
season	Descrição da Época	nvarchar(50) NOT NULL
seasonConceded_overall	Número Total de Golos Sofridos	int NOT NULL
seasonGoalsTotal_overall	Número Total de Golos Marcados e Sofridos	int NOT NULL
seasonScoredNum_overall	Número Total de Golos Marcados	int NOT NULL
seasonConcededNum_overall	Número Total de Golos Sofridos	int NOT NULL

seasonGoalsMin_overall	Minutos por Golo Marcado e Sofrido	int NOT NULL
seasonScoredMin_overall	Minutos por Golo Marcado	int NOT NULL
seasonConcededMin_overall	Minutos por Golo Sofrido	int NOT NULL
seasonGoalDifference_overall	Diferença Golos	int NOT NULL
seasonWinsNum_overall	Número Total de Vitórias	int NOT NULL
seasonDrawsNum_overall	Número Total de Empates	int NOT NULL
seasonLossesNum_overall	Número Total de Derrotas	int NOT NULL
seasonCS_overall	Número Total de Jogos sem Sofrer Golos	int NOT NULL
seasonCSPercentage_overall	Percentagem de Jogos sem Sofrer Golos	int NOT NULL
seasonAVG_overall	Média de Golos por Jogo	decimal (6,3) NOT NULL
seasonScoredAVG_overall	Média de Golos Marcados por Jogo	decimal (6,3) NOT NULL
seasonConcededAVG_overall	Média de Golos Sofridos por Jogo	decimal (6,3) NOT NULL
winPercentage_overall	Percentagem de Vitórias	int NOT NULL
drawPercentage_overall	Percentagem de Empates	int NOT NULL
losePercentage_overall	Percentagem de Derrotas	int NOT NULL
cornersAVG_overall	Média de Cantos a Favor por Jogo	decimal (6,3) NOT NULL
shotsAVG_overall	Média de Remates por Jogo	decimal (6,3) NOT NULL
shotsOnTargetAVG_overall	Média de Remates à Baliza por Jogo	decimal (6,3) NOT NULL
shotsOffTargetAVG_overall	Média de Remates Fora do Alvo por Jogo	decimal (6,3) NOT NULL
possessionAVG_overall	Média de Posse de Bola por Jogo	int NOT NULL
dangerous_attacks_num_overall	Número Total de Ataques Perigosos	int NOT NULL
attacks_num_overall	Número Total de Ataques	int NOT NULL
dangerous_attacks_avg_overall	Média de Ataques Perigosos	decimal (6,3) NOT NULL
attacks_avg_overall	Média de Ataques	decimal (6,3) NOT NULL
cards_for_overall	Número Total de Cartões que a Equipa Recebeu	int
cards_for_avg_overall	Média de Cartões que a Equipa Recebeu	decimal (6,3)
penalties_won_overall	Número de Penaltis Assinalados a Favor	int
penalties_scored_overall	Número de Penaltis Marcados	int
penalties_missed_overall	Número de Penaltis Falhados	int
shots_per_goals_scored_overall	Número de Remates por Golo Marcado	decimal (6,3)
shots_on_target_per_goals_scored_overall	Número de Remates à Baliza por Golo Marcado	decimal (6,3)

shot_conversion_rate_overall	Taxa de Conversão de Remates	decimal (6,3)
------------------------------	------------------------------	---------------

Tabela de Equipas

Atributo	Descrição	Tipo
id	ID da Equipa	int NOT NULL
name	Nome da Equipa (Original)	nvarchar(110) NOT NULL
cleanName	Versão Abreviada do Nome da Equipa	nvarchar(100) NOT NULL
continent	Continente do Jogador	nvarchar(1020)
country	Nome do País (Inglês)	nvarchar(30) NOT NULL
founded	Ano de Fundação do Clube	nvarchar(50) NOT NULL
image	URL da Insígnia da Liga	nvarchar(300) NOT NULL

Anexo 3 – Tabela de Mapeamento Lógico de Dados (Total)

Este anexo mostra a tabela de mapeamento lógico de dados. Esta é a versão completa da que se encontra presente na Secção 3.9.

Destination Table					Source Table				Transformations
Table Name	Column Name	Data Type	Table Type	SCD Type	Database	Table Name	Column Name	Data Type	
DimData	dateKey	int	Dimension	N/A	DataMart				Surrogate Key
DimData	fullDate	datetime	Dimension	N/A	DataMart				
DimData	year	int	Dimension	N/A	DataMart				
DimData	trimester	tinyint	Dimension	N/A	DataMart				
DimData	semester	tinyint	Dimension	N/A	DataMart				
DimData	quarter	tinyint	Dimension	N/A	DataMart				
DimData	month	tinyint	Dimension	N/A	DataMart				
DimData	monthName	nvarchar (10)	Dimension	N/A	DataMart				
DimData	week	tinyint	Dimension	N/A	DataMart				
DimData	dayNumberOfMonth	int	Dimension	N/A	DataMart				
DimData	dayOfWeek	nvarchar (20)	Dimension	N/A	DataMart				

DimData	weekend	nvarchar (3)	Dimension	N/A	DataMart				
DimTempo_Jogo	event_timeKey	int	Dimension	N/A	DataMart				Surrogate Key
DimTempo_Jogo	time	int	Dimension	N/A	DataMart				
DimTempo_Jogo	time_15min	int	Dimension	N/A	DataMart				
DimTempo_Jogo	time_30min	int	Dimension	N/A	DataMart				
DimTempo_Jogo	part	int	Dimension	N/A	DataMart				
DimPais	countryKey	int	Dimension	N/A	Staging Area				Surrogate Key
DimPais	continent_s	nvarchar (10)	Dimension	1	DataMart				SELECT CASE WHEN c.iso IN ('mc','xk','gb','gg','je','li','ch',' rs','sm','mk','fo','me','tr','ba' ,'al','ad','gi','fr','es','pt','it', 'be','nl','lu','de','at','fi', 'se','dk','no','is','ee','lv', 'It','pl','cz','sk','hu','ro', 'hr','si','bg','gr','cy','mt', 'ie','gb-eng','gb-wls','gb- nir','gb-sct','ua','am','md', 'ge','by') THEN 'EU' -- Europe WHEN c.iso IN ('er','cf','cg','yt','td','gq','ss','r

								<p>e','cd','cm','km','dz','eg','ly', 'ma','sd','tn','eh','mr','ml', 'gn','gw','lr','sl','bf','bj', 'ne','tg','gh','ng','ga','gm', 'sn','ci','bf','ng','bj','et', 'so','dj','ke','tz','ug','bi', 'rw','mw','mg','mu','sc', 'zm','zw','mz','ao','na', 'za','bw','sz','ls','st','cv') THEN 'AFR' -- Africa WHEN c.iso IN ('bn','kr','tw','vn','kr','kp','la', 'bt','kh','mm','ph','mn','mo', 'hk','pk','mv','lk','np','tl','th', 'cn','ye','iq','in','my','bd','om', 'sg','sy','kw','jo','ps','qa','ae', 'sa','bh','lb','af','il','ir','id','jp', 'ru','az','kz','kg','tj','tm', 'uz') THEN 'ASIA' -- Asia WHEN c.iso IN ('ca','us', 'mx','bz','gt','sv','hn','ni', 'cr','pa','cu','do','jm','ht', 'bs','bb','tt','kn','lc','vc', 'gd','ag','dm','mq','gp', 'bl','mf','pm','tc','vg','ai', 'bm','ky','vg','ms','tc','bs', 'pr','vi','cu','pa','ht','cw', 'aw','sx') THEN 'NA' -- North America WHEN c.iso IN ('gf','ar', 'bo','br','cl','co','ec','gy',</p>
--	--	--	--	--	--	--	--	--

									<pre>'pe', 'py', 'sr', 'uy', 've') THEN 'SA' -- South America WHEN c.iso IN ('as','au', 'nz', 'fj', 'pg', 'sb', 'vu', 'fm', 'mh', 'pw', 'ki', 'to', 'ws', 'tv', 'mp', 'gu', 'nf', 'nc', 'pf', 'ck', 'nu', 'wf', 'um') THEN 'OC' -- Oceania ELSE 'Desconhecido' END AS continent_s FROM countries c</pre>
DimPais	continent_name	nvarchar (50)	Dimension	1	DataMart				<pre>SELECT CASE WHEN c.iso IN ('mc','xk','gb','gg','je','li','ch', rs','sm','mk','fo','me','tr','ba', 'al','ad','gi','fr', 'es', 'pt', 'it', 'be', 'nl', 'lu', 'de', 'at', 'fi', 'se', 'dk', 'no', 'is', 'ee', 'lv', 'It', 'pl', 'cz', 'sk', 'hu', 'ro', 'hr', 'si', 'bg', 'gr', 'cy', 'mt', 'ie', 'gb-eng', 'gb-wls', 'gb- nir', 'gb-sct','ua','am','md', 'ge', 'by') THEN 'Europa' -- Europe WHEN c.iso IN ('er','cf','cg','yt','td','gq','ss','r e','cd','cm','km','dz', 'eg', 'ly', 'ma', 'sd', 'tn', 'eh', 'mr', 'ml', 'gn', 'gw', 'lr', 'sl', 'bf', 'bj', 'ne', 'tg', 'gh', 'ng', 'ga', 'gm',</pre>

								'sn', 'ci', 'bf', 'ng', 'bj', 'et', 'so', 'dj', 'ke', 'tz', 'ug', 'bi', 'rw', 'mw', 'mg', 'mu', 'sc', 'zm', 'zw', 'mz', 'ao', 'na', 'za', 'bw', 'sz', 'ls', 'st', 'cv') THEN 'África' -- Africa WHEN c.iso IN ('bn', 'kr', 'tw', 'vn', 'kr', 'kp', 'la', 'bt', 'kh', 'mm', 'ph', 'mn', 'mo', 'hk', 'pk', 'mv', 'lk', 'np', 'tl', 'th', 'cn', 'ye', 'iq', 'in', 'my', 'bd', 'om', ',sg', 'sy', 'kw', 'jo', 'ps', 'qa', 'ae', ',sa', 'bh', 'lb', 'af', 'il', 'ir', 'id', 'jp', ',ru', 'az', 'kz', 'kg', 'tj', 'tm', 'uz') THEN 'Ásia' -- Asia WHEN c.iso IN ('ca', 'us', 'mx', 'bz', 'gt', 'sv', 'hn', 'ni', 'cr', 'pa', 'cu', 'do', 'jm', 'ht', 'bs', 'bb', 'tt', 'kn', 'lc', 'vc', 'gd', 'ag', 'dm', 'mq', 'gp', 'bl', 'mf', 'pm', 'tc', 'vg', 'ai', 'bm', 'ky', 'vg', 'ms', 'tc', 'bs', 'pr', 'vi', 'cu', 'pa', 'ht', 'cw', 'aw', 'sx') THEN 'América do Norte' -- North America WHEN c.iso IN ('gf', 'ar', 'bo', 'br', 'cl', 'co', 'ec', 'gy', 'pe', 'py', 'sr', 'uy', 've') THEN 'América do Sul' -- South America WHEN c.iso IN ('as', 'au',
--	--	--	--	--	--	--	--	--

									'nz', 'fj', 'pg', 'sb', 'vu', 'fm', 'mh', 'pw', 'ki', 'to', 'ws', 'tv', 'mp', 'gu', 'nf', 'nc', 'pf', 'ck', 'nu', 'wf', 'um') THEN 'Oceania' -- Oceania ELSE 'Desconhecido' END AS continent_name FROM countries c
DimPais	country_ID	int	Dimension	1	Staging Area	Pais	country_ID	int	Chave do Sistema Operacional
DimPais	iso	nvarchar (80)	Dimension	1	Staging Area	Pais	iso	nvarchar (80)	
DimPais	iso_number	int	Dimension	1	Staging Area	Pais	iso_number	int	
DimPais	country	nvarchar (120)	Dimension	1	Staging Area	Pais	country	nvarchar (120)	
DimPais	name_pt	nvarchar (120)	Dimension	1	Staging Area	Pais	name_pt	nvarchar (120)	
DimLiga	competitionKey	int	Dimension	N/A	Staging Area				Surrogate Key
DimLiga	id	int	Dimension	1	Staging Area	Liga	id	int	Chave do Sistema Operacional
DimLiga	name	nvarchar (40)	Dimension	1	Staging Area	Liga	name	nvarchar (40)	
DimLiga	english_name	nvarchar (40)	Dimension	1	Staging Area	Liga	english_name	nvarchar (40)	
DimLiga	domestic_scale	int	Dimension	1	Staging Area	Liga	domestic_scale	int	
DimLiga	international_scale	int	Dimension	2	Staging Area	Liga	international_scale	int	

DimLiga	status	nvarchar (50)	Dimension	1	Staging Area	Liga	status	nvarchar (50)	
DimLiga	start_yearKey	int	Dimension	N/A	Staging Area, Data Mart	DimYear	yearKey	int	where SA_BI_FOOTBALL_1210129.Leagues.starting_year = DM_BI_FOOTBALL_1210129.DimYear.year
DimLiga	end_yearKey	int	Dimension	N/A	Staging Area, Data Mart	DimYear	yearKey	int	where SA_BI_FOOTBALL_1210129.Leagues.ending_year = DM_BI_FOOTBALL_1210129.DimYear.year
DimLiga	image	nvarchar (260)	Dimension	1	Staging Area	Liga	image	nvarchar (260)	
DimLiga	clubNum	int	Dimension	1	Staging Area	Liga	clubNum	int	
DimLiga	season	nvarchar (36)	Dimension	1	Staging Area	Liga	season	nvarchar (36)	
DimLiga	effectiveDate	date	Dimension	N/A					
DimLiga	expiredDate	date	Dimension	N/A					
DimLiga	isCurrent	nvarchar (3)	Dimension	N/A					
FactJogo	matchKey	int	Fact	N/A	Staging Area, Data Mart	DimMatch	teamKey	int	Foreign Key where SA_BI_FOOTBALL_1210129.Matches.ID = DM_BI_FOOTBALL_1210129.DimMatch.id
FactJogo	homeKey	int	Fact	N/A	Staging Area,	DimTeam	teamKey	int	Primary Key

					Data Mart				where SA_BI_FOOTBALL_1210129. Matches.homeID = DM_BI_FOOTBALL_1210129 .DimTeam.id
FactJogo	awayKey	int	Fact	N/A	Staging Area, Data Mart	DimTeam	teamKey	int	Primary Key where SA_BI_FOOTBALL_1210129. Matches.awayID = DM_BI_FOOTBALL_1210129 .DimTeam.id
FactJogo	stadiumKey	int	Fact	N/A	Staging Area, Data Mart	DimStadium	stadiumKey	int	Primary Key where SA_BI_FOOTBALL_1210129. Matches.stadium_name = DM_BI_FOOTBALL_1210129 .DimLeague.stadium_name
FactJogo	competitionKey	int	Fact	N/A	Staging Area, Data Mart	DimLeague	competitionKey	int	Primary Key where SA_BI_FOOTBALL_1210129. Matches.competition_id = DM_BI_FOOTBALL_1210129 .DimLeague.id
FactJogo	id	int	Fact	N/A	Staging Area	Jogo	id	int	Chave do Sistema Operacional
FactJogo	homeGoalCount	int	Fact	N/A	Staging Area	Jogo	homeGoalCount	int	

FactJogo	awayGoalCount	int	Fact	N/A	Staging Area	Jogo	awayGoalCount	int	
FactJogo	totalGoalCount	int	Fact	N/A	Staging Area	Jogo	totalGoalCount	int	
FactJogo	team_a_yellow_cards	int	Fact	N/A	Staging Area	Jogo	team_a_yellow_cards	int	
FactJogo	team_b_yellow_cards	int	Fact	N/A	Staging Area	Jogo	team_b_yellow_cards	int	
FactJogo	team_a_red_cards	int	Fact	N/A	Staging Area	Jogo	team_a_red_cards	int	
FactJogo	team_b_red_cards	int	Fact	N/A	Staging Area	Jogo	team_b_red_cards	int	
FactJogo	team_a_shotsOnTarget	int	Fact	N/A	Staging Area	Jogo	team_a_shotsOnTarget	int	
FactJogo	team_b_shotsOnTarget	int	Fact	N/A	Staging Area	Jogo	team_b_shotsOnTarget	int	
FactJogo	team_a_shotsOffTarget	int	Fact	N/A	Staging Area	Jogo	team_a_shotsOffTarget	int	
FactJogo	team_b_shotsOffTarget	int	Fact	N/A	Staging Area	Jogo	team_b_shotsOffTarget	int	
FactJogo	team_a_shots	int	Fact	N/A	Staging Area	Jogo	team_a_shots	int	
FactJogo	team_b_shots	int	Fact	N/A	Staging Area	Jogo	team_b_shots	int	

FactJogo	team_a_fouls	int	Fact	N/A	Staging Area	Jogo	team_a_fouls	int	
FactJogo	team_b_fouls	int	Fact	N/A	Staging Area	Jogo	team_b_fouls	int	
FactJogo	team_a_possession	int	Fact	N/A	Staging Area	Jogo	team_a_possession	int	
FactJogo	team_b_possession	int	Fact	N/A	Staging Area	Jogo	team_b_possession	int	
FactJogo	team_a_cards_num	int	Fact	N/A	Staging Area	Jogo	team_a_cards_num	int	
FactJogo	team_b_cards_num	int	Fact	N/A	Staging Area	Jogo	team_b_cards_num	int	
FactJogo	dateKey	int	Dimension	N/A	Staging Area, Data Mart	DimDate	dateKey	int	Primary Key where CONVERT(DATE, DATEADD(SECOND, (SA_BI_FOOTBALL_1210129.Matches.date_unix), '19700101')) = DM_BI_FOOTBALL_1210129.DimDate.fullDate
FactJogo	team_a_dangerous_attacks	int	Fact	N/A	Staging Area	Jogo	team_a_dangerous_attacks	int	
FactJogo	team_b_dangerous_attacks	int	Fact	N/A	Staging Area	Jogo	team_b_dangerous_attacks	int	
FactJogo	team_a_attacks	int	Fact	N/A	Staging Area	Jogo	team_a_attacks	int	

FactJogo	team_b_attacks	int	Fact	N/A	Staging Area	Jogo	team_b_attacks	int	
FactJogo	team_a_penalties_won	int	Fact	N/A	Staging Area	Jogo	team_a_penalties_won	int	
FactJogo	team_b_penalties_won	int	Fact	N/A	Staging Area	Jogo	team_b_penalties_won	int	
FactJogo	team_a_penalty_goals	int	Fact	N/A	Staging Area	Jogo	team_a_penalty_goals	int	
FactJogo	team_b_penalty_goals	int	Fact	N/A	Staging Area	Jogo	team_b_penalty_goals	int	
FactJogo	team_a_penalty_missed	int	Fact	N/A	Staging Area	Jogo	team_a_penalty_missed	int	
FactJogo	team_b_penalty_missed	int	Fact	N/A	Staging Area	Jogo	team_b_penalty_missed	int	
FactJogador_Jogo	playerKey	int	Fact	N/A	Staging Area, Data Mart	DimJogador	playerKey	int	Primary Key where SA_BI_FOOTBALL_1210129. MatchesPlayers.playerID = DM_BI_FOOTBALL_1210129. .DimPlayer.id
FactJogador_Jogo	eventKey	int	Fact	N/A	Staging Area, Data Mart	DimEvent_Match	eventKey	int	Primary Key where SA_BI_FOOTBALL_1210129.

									MatchesPlayers.event_type = DM_BI_FOOTBALL_1210129.DimEvent_Match.event_type
FactJogador_Jogo	event_timeKey	int	Fact	1	Staging Area	Jogador_Jogo	event_time	nvarchar (20)	Primary Key SELECT CASE WHEN CHARINDEX('+', event_time) > 0 THEN CAST(SUBSTRING(event_time, 1, CHARINDEX('+', event_time) - 1) AS INT) + CAST(SUBSTRING(event_time, CHARINDEX('+', event_time) + 1, LEN(event_time) - CHARINDEX('+', event_time) - 1) AS INT) ELSE CAST(SUBSTRING(event_time, 1, LEN(event_time) - 1) AS INT) END AS event_time FROM MatchesPlayers
FactJogador_Jogo	matchKey	int	Fact	N/A	Staging Area, Data Mart	FactMatches	id	int	Primary Key

									where SA_BI_FOOTBALL_1210129. MatchesPlayers.matchID = SA_BI_FOOTBALL_1210129. FactMatches.id
FactJogador_Jogo	competitionKey	int	Fact	N/A	Staging Area, Data Mart	DimLeague	competitionKey	int	Foreign Key where SA_BI_FOOTBALL_1210129. MatchesPlayers.competitionID = DM_BI_FOOTBALL_1210129. .DimLeague.id
FactJogador_Jogo	stadiumKey	int	Fact	N/A	Staging Area, Data Mart	DimStadium	stadiumKey	int	Foreign Key where SA_BI_FOOTBALL_1210129. MatchesPlayers.stadium_name = DM_BI_FOOTBALL_1210129. .DimLeague.stadium_name
FactJogador_Jogo	awayKey	int	Fact	N/A	Staging Area, Data Mart	DimTeam	teamKey	int	Foreign Key where SA_BI_FOOTBALL_1210129. MatchesPlayers.awayID = DM_BI_FOOTBALL_1210129. .DimTeam.id
FactJogador_Jogo	shirt_number	int	Fact	N/A	Staging Area	Jogador_Jogo	shirt_number	int	

FactJogador_Jogo	player_outKey	int	Fact	N/A	Staging Area, Data Mart	DimJogador	playerKey	int	Foreign Key where SA_BI_FOOTBALL_1210129. MatchesPlayers.player_out_id = DM_BI_FOOTBALL_1210129. DimPlayer.id
FactJogador_Jogo	player_out_time Key	int	Fact	N/A	Staging Area	Jogador_Jogo	player_out_time	nvarchar (20)	Foreign Key SELECT CASE WHEN CHARINDEX('+', player_out_time) > 0 THEN CAST(SUBSTRING(player_out_time, 1, CHARINDEX('+', player_out_time) - 1) AS INT) + CAST(SUBSTRING(player_out_time, CHARINDEX('+', player_out_time) + 1, LEN(player_out_time) - CHARINDEX('+', player_out_time) - 1) AS INT) ELSE CAST(SUBSTRING(player_out_time, 1, LEN(player_out_time) - 1) AS INT)

									END AS player_out_time FROM MatchesPlayers
DimEvento_Jogo	eventKey	int	Fact	N/A					Surrogate Key
DimEvento_Jogo	event_type	nvarchar (50)	Dimension	1	Staging Area	Jogad or_Jog o	event_typ e	nvarchar (30)	SELECT distinct (event_type) FROM MatchesPlayers
DimEvento_Jogo	bench	nvarchar (3)	Dimension	1	Staging Area	Jogad or_Jog o	bench	bit	
DimEvento_Jogo	lineup	nvarchar (3)	Dimension	1	Staging Area	Jogad or_Jog o	lineup	bit	
DimEvento_Jogo	home	nvarchar (3)	Dimension	1	Staging Area	Jogad or_Jog o	home	bit	
DimEvento_Jogo	away	nvarchar (3)	Dimension	1	Staging Area	Jogad or_Jog o	away	bit	
DimJogo	matchKey	int	Dimension	N/A					Surrogate Key
DimJogo	id	int	Fact	N/A	Staging Area	Jogo	id	int	Chave do Sistema Operacional
DimJogo	match_descripti on	nvarchar (350)	Dimension	1	Staging Area	Jogo, Equip as	cleanNam e, stadium_n ame,	nvarchar (350)	SELECT distinct (B.cleanName + ' - ' + C.cleanName + ' (' + stadium_name + ')') As match_description FROM SA_BI_FOOTBALL_1210129. dbo.Matches A inner join SA_BI_FOOTBALL_1210129.

									dbo.Teams B on A.homeID=B.id inner join SA_BI_FOOTBALL_1210129. dbo.Teams C on A.awayID=C.id
FactJogador_Liga	playerKey	int	Fact	N/A	Staging Area, Data Mart	DimPl ayer	playerKey	int	Primary Key where SA_BI_FOOTBALL_1210129. LeaguePlayers.player_id = DM_BI_FOOTBALL_1210129 .DimPlayer.id
FactJogador_Liga	competitionKey	int	Fact	N/A	Staging Area, Data Mart	DimLe age	competitio nKey	int	Primary Key where SA_BI_FOOTBALL_1210129. LeaguePlayers.competition_ id = DM_BI_FOOTBALL_1210129 .DimLeague.id
FactJogador_Liga	teamKey	int	Fact	N/A	Staging Area, Data Mart	DimTe am	teamKey	int	Primary Key where SA_BI_FOOTBALL_1210129. LeaguePlayers.club_team_i d = DM_BI_FOOTBALL_1210129 .DimTeam.id
FactJogador_Liga	start_yearKey	int	Fact	N/A	Staging Area,	DimYe ar	yearKey	int	Primary Key

					Data Mart				where SA_BI_FOOTBALL_1210129. LeaguePlayers.starting_year = DM_BI_FOOTBALL_1210129 .DimYear.year
FactJogador_Liga	end_yearKey	int	Fact	N/A	Staging Area, Data Mart	DimYear	yearKey	int	Primary Key where SA_BI_FOOTBALL_1210129. LeaguePlayers.ending_year = DM_BI_FOOTBALL_1210129 .DimYear.year
FactJogador_Liga	team_2Key	int	Fact	N/A	Staging Area, Data Mart	DimTeam	teamKey	int	Foreign Key where SA_BI_FOOTBALL_1210129. LeaguePlayers.club_team_2 _id = DM_BI_FOOTBALL_1210129 .DimTeam.id
FactJogador_Liga	minutes_played_overall	int	Fact	N/A	Staging Area	Jogador_Liga	minutes_played_overall	int	
FactJogador_Liga	appearances_overall	int	Fact	N/A	Staging Area	Jogador_Liga	appearances_overall	int	
FactJogador_Liga	goals_overall	int	Fact	N/A	Staging Area	Jogador_Liga	goals_overall	int	

FactJogador_Liga	clean_sheets_overall	int	Fact	N/A	Staging Area	Jogador_Liga	clean_sheets_overall	int	
FactJogador_Liga	conceded_overall	int	Fact	N/A	Staging Area	Jogador_Liga	conceded_overall	int	
FactJogador_Liga	assists_overall	int	Fact	N/A	Staging Area	Jogador_Liga	assists_overall	int	
FactJogador_Liga	penalty_goals	int	Fact	N/A	Staging Area	Jogador_Liga	penalty_goals	int	
FactJogador_Liga	penalty_misses	int	Fact	N/A	Staging Area	Jogador_Liga	penalty_misses	int	
FactJogador_Liga	penalty_success	int	Fact	N/A	Staging Area	Jogador_Liga	penalty_success	int	
FactJogador_Liga	goals_involved_per_90_overall	decimal (6,3)	Fact	N/A	Staging Area	Jogador_Liga	goals_involved_per_90_overall	decimal (6,3)	
FactJogador_Liga	assists_per_90_overall	decimal (6,3)	Fact	N/A	Staging Area	Jogador_Liga	assists_per_90_overall	decimal (6,3)	
FactJogador_Liga	goals_per_90_overall	decimal (6,3)	Fact	N/A	Staging Area	Jogador_Liga	goals_per_90_overall	decimal (6,3)	
FactJogador_Liga	clean_sheets_per_overall	int	Fact	N/A	Staging Area	Jogador_Liga	clean_sheets_per_overall	int	

FactJogador_Liga	min_per_goal_overall	int	Fact	N/A	Staging Area	Jogador_Liga	min_per_goal_overall	int	
FactJogador_Liga	conceded_per_90_overall	decimal (6,3)	Fact	N/A	Staging Area	Jogador_Liga	conceded_per_90_overall	decimal (6,3)	
FactJogador_Liga	min_per_conceded_overall	int	Fact	N/A	Staging Area	Jogador_Liga	min_per_conceded_overall	int	
FactJogador_Liga	cards_overall	int	Fact	N/A	Staging Area	Jogador_Liga	cards_overall	int	
FactJogador_Liga	yellow_cards_overall	int	Fact	N/A	Staging Area	Jogador_Liga	yellow_cards_overall	int	
FactJogador_Liga	red_cards_overall	int	Fact	N/A	Staging Area	Jogador_Liga	red_cards_overall	int	
FactJogador_Liga	min_per_match	int	Fact	N/A	Staging Area	Jogador_Liga	min_per_match	int	
FactJogador_Liga	min_per_card_overall	int	Fact	N/A	Staging Area	Jogador_Liga	min_per_card_overall	int	
FactJogador_Liga	min_per_assist_overall	int	Fact	N/A	Staging Area	Jogador_Liga	min_per_assist_overall	int	
FactJogador_Liga	cards_per_90_overall	decimal (6,3)	Fact	N/A	Staging Area	Jogador_Liga	cards_per_90_overall	decimal (6,3)	

FactJogador_Liga	rank_in_league_top_attackers	int	Fact	N/A	Staging Area	Jogador_Liga	rank_in_league_top_attackers	int	
FactJogador_Liga	rank_in_league_top_midfielders	int	Fact	N/A	Staging Area	Jogador_Liga	rank_in_league_top_midfielders	int	
FactJogador_Liga	rank_in_league_top_defenders	int	Fact	N/A	Staging Area	Jogador_Liga	rank_in_league_top_defenders	int	
FactJogador_Liga	rank_in_club_top_scorer	int	Fact	N/A	Staging Area	Jogador_Liga	rank_in_club_top_scorer	int	
FactJogador_Liga	last_matchKey	int	Fact	N/A	Staging Area, Data Mart	DimDate	dateKey	int	Foreign Key where CONVERT(DATE, DATEADD(SECOND, (SA_BI_FOOTBALL_1210129.LigaPlayers.last_match_timestamp), '19700101')) = DM_BI_FOOTBALL_1210129.DimDate.fullDate
DimJogador	playerKey	int	Dimension	N/A					Surrogate Key
DimJogador	id	int	Dimension	1	Staging Area	Jogador	id	int	
DimJogador	full_name	nvarchar (160)	Dimension	1	Staging Area	Jogador	full_name	nvarchar (160)	
DimJogador	first_name	nvarchar (80)	Dimension	1	Staging Area	Jogador	first_name	nvarchar (80)	

DimJogador	last_name	nvarchar (120)	Dimension	1	Staging Area	Jogador	last_name	nvarchar (120)	
DimJogador	known_as	nvarchar (140)	Dimension	2	Staging Area	Jogador	known_as	nvarchar (140)	
DimJogador	height	int	Dimension	1	Staging Area	Jogador	height	int	
DimJogador	weight	int	Dimension	1	Staging Area	Jogador	weight	int	
DimJogador	position	nvarchar (40)	Dimension	2	Staging Area	Jogador	position	nvarchar (40)	
DimJogador	age	int	Dimension	1	Staging Area	Jogador	birthday	int	SELECT DATEDIFF(YEAR, CONVERT(DATE, DATEADD(SECOND, birthday, '19700101')), convert(date,getdate())) AS age FROM Players
DimJogador	countryKey	int	Dimension	N/A	Staging Area, Data Mart	DimCountry	countryKey	int	Foreign Key where SA_BI_FOOTBALL_1210129.Players.nationality = DM_BI_FOOTBALL_1210129.DimCountry.country
DimJogador	birthdayKey	int	Dimension	N/A	Staging Area, Data Mart	DimDate	yearKey	int	Foreign Key where CONVERT(DATE, DATEADD(SECOND, (SA_BI_FOOTBALL_1210129.Players.birthday), '19700101')) =

									DM_BI_FOOTBALL_1210129 .DimDate.fullDate
DimJogador	effectiveDate	date	Dimension	N/A					
DimJogador	expiredDate	date	Dimension	N/A					
DimJogador	isCurrent	nvarchar (3)	Dimension	N/A					
DimEstadio	stadiumKey	int	Dimension	N/A					Surrogate Key
DimEstadio	stadium_addresses	nvarchar (110)	Dimension	1	Staging Area	Estadio	stadium_address	nvarchar (110)	
DimEstadio	stadium_name	nvarchar (60)	Dimension	2	Staging Area	Estadio	stadium_name	nvarchar (60)	
DimEstadio	effectiveDate	date	Dimension	N/A					
DimEstadio	expiredDate	date	Dimension	N/A					
DimEstadio	isCurrent	nvarchar (3)	Dimension	N/A					
FactEquipa_Liga	teamKey	int	Fact	N/A					
FactEquipa_Liga	competitionKey	int	Fact	N/A					
FactEquipa_Liga	start_yearKey	int	Fact	N/A	Staging Area, Data Mart	DimTeam	teamKey	int	Primary Key where SA_BI_FOOTBALL_1210129. LeagueTeams.team_id = DM_BI_FOOTBALL_1210129 .DimTeam.id
FactEquipa_Liga	end_yearKey	int	Fact	N/A	Staging Area, Data Mart	DimLeague	competitionKey	int	Primary Key where SA_BI_FOOTBALL_1210129. LeagueTeams.competition_id =

									DM_BI_FOOTBALL_1210129 .DimLeague.id
FactEquipa_Liga	table_position	int	Fact	N/A	Staging Area	Equip a_Liga	table_position	int	
FactEquipa_Liga	seasonConceded_overall	int	Fact	N/A	Staging Area	Equip a_Liga	seasonConceded_overall	int	
FactEquipa_Liga	seasonGoalsTotal_overall	int	Fact	N/A	Staging Area	Equip a_Liga	seasonGoalsTotal_overall	int	
FactEquipa_Liga	seasonScoredNum_overall	int	Fact	N/A	Staging Area	Equip a_Liga	seasonScoredNum_overall	int	
FactEquipa_Liga	seasonConcededNum_overall	int	Fact	N/A	Staging Area	Equip a_Liga	seasonConcededNum_overall	int	
FactEquipa_Liga	seasonGoalsMin_overall	int	Fact	N/A	Staging Area	Equip a_Liga	seasonGoalsMin_overall	int	
FactEquipa_Liga	seasonScoredMin_overall	int	Fact	N/A	Staging Area	Equip a_Liga	seasonScoredMin_overall	int	
FactEquipa_Liga	seasonConcededMin_overall	int	Fact	N/A	Staging Area	Equip a_Liga	seasonConcededMin_overall	int	
FactEquipa_Liga	seasonGoalDifference_overall	int	Fact	N/A	Staging Area	Equip a_Liga	seasonGoalDifference_overall	int	
FactEquipa_Liga	seasonWinsNum_overall	int	Fact	N/A	Staging Area	Equip a_Liga	seasonWinsNum_overall	int	

FactEquipa_Liga	seasonDrawsNum_overall	int	Fact	N/A	Staging Area	Equipa_Liga	seasonDrawsNum_overall	int	
FactEquipa_Liga	seasonLossesNum_overall	int	Fact	N/A	Staging Area	Equipa_Liga	seasonLossesNum_overall	int	
FactEquipa_Liga	seasonCS_overall	int	Fact	N/A	Staging Area	Equipa_Liga	seasonCS_overall	int	
FactEquipa_Liga	seasonCSPercentage_overall	int	Fact	N/A	Staging Area	Equipa_Liga	seasonCSPercentage_overall	int	
FactEquipa_Liga	seasonAVG_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	seasonAVG_overall	decimal (6,3)	
FactEquipa_Liga	seasonScoredAVG_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	seasonScoredAVG_overall	decimal (6,3)	
FactEquipa_Liga	seasonConcededAVG_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	seasonConcededAVG_overall	decimal (6,3)	
FactEquipa_Liga	winPercentage_overall	int	Fact	N/A	Staging Area	Equipa_Liga	winPercentage_overall	int	
FactEquipa_Liga	drawPercentage_overall	int	Fact	N/A	Staging Area	Equipa_Liga	drawPercentage_overall	int	
FactEquipa_Liga	losePercentage_overall	int	Fact	N/A	Staging Area	Equipa_Liga	losePercentage_overall	int	
FactEquipa_Liga	shotsAVG_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	shotsAVG_overall	decimal (6,3)	

FactEquipa_Liga	shotsOnTargetAVG_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	shotsOnTargetAVG_overall	decimal (6,3)	
FactEquipa_Liga	shotsOffTargetAVG_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	shotsOffTargetAVG_overall	decimal (6,3)	
FactEquipa_Liga	possessionAVG_overall	int	Fact	N/A	Staging Area	Equipa_Liga	possessionAVG_overall	int	
FactEquipa_Liga	dangerous_attacks_num_overall	int	Fact	N/A	Staging Area	Equipa_Liga	dangerous_attacks_num_overall	int	
FactEquipa_Liga	attacks_num_overall	int	Fact	N/A	Staging Area	Equipa_Liga	attacks_num_overall	int	
FactEquipa_Liga	dangerous_attacks_avg_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	dangerous_attacks_avg_overall	decimal (6,3)	
FactEquipa_Liga	attacks_avg_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	attacks_avg_overall	decimal (6,3)	
FactEquipa_Liga	cards_for_overall	int	Fact	N/A	Staging Area	Equipa_Liga	cards_for_overall	int	
FactEquipa_Liga	cards_for_avg_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	cards_for_avg_overall	decimal (6,3)	
FactEquipa_Liga	penalties_won_overall	int	Fact	N/A	Staging Area	Equipa_Liga	penalties_won_overall	int	
FactEquipa_Liga	penalties_scored_overall	int	Fact	N/A	Staging Area	Equipa_Liga	penalties_scored_overall	int	

FactEquipa_Liga	penalties_missed_overall	int	Fact	N/A	Staging Area	Equipa_Liga	penalties_missed_overall	int	
FactEquipa_Liga	shots_per_goals_scored_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	shots_per_goals_scored_overall	decimal (6,3)	
FactEquipa_Liga	shots_on_target_per_goals_scored_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	shots_on_target_per_goals_scored_overall	decimal (6,3)	
FactEquipa_Liga	shot_conversion_rate_overall	decimal (6,3)	Fact	N/A	Staging Area	Equipa_Liga	shot_conversion_rate_overall	decimal (6,3)	
FactEquipa_Liga	goals_all_min_0_to_15	int	Fact	N/A	Staging Area	Equipa_Liga	goals_all_min_0_to_15	int	
FactEquipa_Liga	goals_all_min_16_to_30	int	Fact	N/A	Staging Area	Equipa_Liga	goals_all_min_16_to_30	int	
FactEquipa_Liga	goals_all_min_31_to_45	int	Fact	N/A	Staging Area	Equipa_Liga	goals_all_min_31_to_45	int	
FactEquipa_Liga	goals_all_min_46_to_60	int	Fact	N/A	Staging Area	Equipa_Liga	goals_all_min_46_to_60	int	
FactEquipa_Liga	goals_all_min_61_to_75	int	Fact	N/A	Staging Area	Equipa_Liga	goals_all_min_61_to_75	int	

FactEquipa_Liga	goals_all_min_76_to_90	int	Fact	N/A	Staging Area	Equip a_Liga	goals_all_min_76_to_90	int	
FactEquipa_Liga	goals_scored_min_0_to_15	int	Fact	N/A	Staging Area	Equip a_Liga	goals_scored_min_0_to_15	int	
FactEquipa_Liga	goals_scored_min_16_to_30	int	Fact	N/A	Staging Area	Equip a_Liga	goals_scored_min_16_to_30	int	
FactEquipa_Liga	goals_scored_min_31_to_45	int	Fact	N/A	Staging Area	Equip a_Liga	goals_scored_min_31_to_45	int	
FactEquipa_Liga	goals_scored_min_46_to_60	int	Fact	N/A	Staging Area	Equip a_Liga	goals_scored_min_46_to_60	int	
FactEquipa_Liga	goals_scored_min_61_to_75	int	Fact	N/A	Staging Area	Equip a_Liga	goals_scored_min_61_to_75	int	
FactEquipa_Liga	goals_scored_min_76_to_90	int	Fact	N/A	Staging Area	Equip a_Liga	goals_scored_min_76_to_90	int	
FactEquipa_Liga	goals_conceded_min_0_to_15	int	Fact	N/A	Staging Area	Equip a_Liga	goals_conceded_min_0_to_15	int	
FactEquipa_Liga	goals_conceded_min_16_to_30	int	Fact	N/A	Staging Area	Equip a_Liga	goals_conceded_min_16_to_30	int	
FactEquipa_Liga	goals_conceded_min_31_to_45	int	Fact	N/A	Staging Area	Equip a_Liga	goals_conceded_min_31_to_45	int	

FactEquipa_Liga	goals_conceded_min_46_to_60	int	Fact	N/A	Staging Area	Equip a_Liga	goals_conceded_min_46_to_60	int	
FactEquipa_Liga	goals_conceded_min_61_to_75	int	Fact	N/A	Staging Area	Equip a_Liga	goals_conceded_min_61_to_75	int	
FactEquipa_Liga	goals_conceded_min_76_to_90	int	Fact	N/A	Staging Area	Equip a_Liga	goals_conceded_min_76_to_90	int	
DimEquipa	teamKey	int	Dimension	N/A					Surrogate Key
DimEquipa	id	int	Dimension	1	Staging Area	Equip a	id	int	
DimEquipa	name	nvarchar(110)	Dimension	1	Staging Area	Equip a	name	nvarchar(110)	
DimEquipa	cleanName	nvarchar(100)	Dimension	1	Staging Area	Equip a	cleanName	nvarchar(100)	
DimEquipa	age	int	Dimension	1	Staging Area	Equip a	founded	nvarchar(50)	SELECT DATEDIFF(YEAR, convert(date,founded), convert(date,getdate())) AS founded FROM Teams
DimEquipa	foundedKey	int	Dimension	N/A	Staging Area, Data Mart	DimDate	yearKey	int	Foreign Key where SA_BI_FOOTBALL_1210129. Teams.founded= DM_BI_FOOTBALL_1210129. .DimDate.fullDate
DimEquipa	image	nvarchar(300)	Dimension	1	Staging Area	Equip a	image	nvarchar(300)	

Anexo 4 – *Scripts* das tabelas da Staging Area

Este anexo exhibe os *scripts* SQL utilizados na criação da *staging area* da solução desenvolvida, descrita na secção 4.2.

Continentes

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'Continents')
CREATE TABLE [dbo].[Continents](
    continent [nvarchar](10) NOT NULL
)
Else
    Truncate TABLE Continents
```

Países

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'Countries')
CREATE TABLE [dbo].[Countries](
    country_ID [int],
    iso [nvarchar](80),
    iso_number int,
    country [nvarchar](120) NOT NULL,
    name_pt [nvarchar](120)
)
Else
    Truncate TABLE Countries
```

Ligas

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'Leagues')
CREATE TABLE [dbo].[Leagues](
    id [int] NOT NULL,
    name [nvarchar](40) NOT NULL,
    english_name [nvarchar](40) NOT NULL,
    country [nvarchar](40) NOT NULL,
    domestic_scale [int] NOT NULL,
```

```

international_scale [int] NOT NULL,
status [nvarchar](50) NOT NULL,
starting_year [int] NOT NULL,
ending_year [int] NOT NULL,
continent [nvarchar](10) NOT NULL,
image [nvarchar](260) NOT NULL,
clubNum[int] NOT NULL,
season [nvarchar](36) NOT NULL
)
Else

```

```
Truncate TABLE Leagues
```

Jogos

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'Matches')
```

```

CREATE TABLE [dbo].[Matches](
    id [int] NOT NULL,
    homeID [int] NOT NULL,
    awayID [int] NOT NULL,
    status [nvarchar](40) NOT NULL,
    roundID[int] NOT NULL,
    homeGoalCount [int] NOT NULL,
    awayGoalCount [int] NOT NULL,
    totalGoalCount [int] NOT NULL,
    team_a_yellow_cards [int] NOT NULL,
    team_b_yellow_cards [int] NOT NULL,
    team_a_red_cards [int] NOT NULL,
    team_b_red_cards [int] NOT NULL,
    team_a_shotsOnTarget [int] NOT NULL,
    team_b_shotsOnTarget [int] NOT NULL,
    team_a_shotsOffTarget [int] NOT NULL,
    team_b_shotsOffTarget [int] NOT NULL,
    team_a_shots [int] NOT NULL,
    team_b_shots [int] NOT NULL,

```

```

team_a_fouls [int] NOT NULL,
team_b_fouls [int] NOT NULL,
team_a_possession [int] NOT NULL,
team_b_possession [int] NOT NULL,
refereeID [int],
coach_a_ID [int],
coach_b_ID [int],
stadium_name [nvarchar](240) NOT NULL,
team_a_cards_num [int] NOT NULL,
team_b_cards_num [int] NOT NULL,
date_unix [int] NOT NULL,
team_a_dangerous_attacks [int] NOT NULL,
team_b_dangerous_attacks [int] NOT NULL,
team_a_attacks [int] NOT NULL,
team_b_attacks [int] NOT NULL,
team_a_penalties_won [int] NOT NULL,
team_b_penalties_won [int] NOT NULL,
team_a_penalty_goals [int] NOT NULL,
team_b_penalty_goals [int] NOT NULL,
team_a_penalty_missed [int] NOT NULL,
team_b_penalty_missed [int] NOT NULL,
competition_id [int] NOT NULL
)
Else
Truncate TABLE Matches

Jogadores

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'Players')
CREATE TABLE [dbo].[Players](
id [int] NOT NULL,
full_name [nvarchar](160) NOT NULL,

```

```

        first_name    [nvarchar](80) NOT NULL,
        last_name     [nvarchar](120) NOT NULL,
        known_as      [nvarchar](140) NOT NULL,
        height [int] NOT NULL,
        weight [int] NOT NULL,
        position      [nvarchar](40) NOT NULL,
        birthday      [int] NOT NULL,
        nationality    [nvarchar](90) NOT NULL,
        continent     [nvarchar](10) NOT NULL
    )

```

Else

```
Truncate TABLE Players
```

Jornadas

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'Rounds')
```

```
CREATE TABLE [dbo].[Rounds](
    roundID [int] NOT NULL,
    awayID [int] NOT NULL,
    homeID [int] NOT NULL,
    game_week [int] NOT NULL,
    competition_id [int] NOT NULL
)

```

Else

```
Truncate TABLE Rounds
```

Estádios

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'Stadiums')
```

```
CREATE TABLE [dbo].[Stadiums](
    stadium_address [nvarchar](110),
    stadium_name [nvarchar](60)
)

```

Else

```
Truncate TABLE Stadiums
```

Equipas

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'Teams')
```

```
CREATE TABLE [dbo].[Teams](  
    id [int] NOT NULL,  
    name [nvarchar](110) NOT NULL,  
    cleanName [nvarchar](100) NOT NULL,  
    continent [nvarchar](1020),  
    country [nvarchar](30) NOT NULL,  
    founded [nvarchar](50) NOT NULL,  
    image [nvarchar](300) NOT NULL  
)
```

```
Else
```

```
Truncate TABLE Teams
```

Jogadores_Jogos

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'MatchesPlayers')
```

```
CREATE TABLE [dbo].[MatchesPlayers](  
    matchID [int] NOT NULL,  
    homeID [int] NOT NULL,  
    awayID [int] NOT NULL,  
    roundID [int] NOT NULL,  
    refereeID [int],  
    stadium_name [nvarchar](250),  
    competitionID [int] NOT NULL,  
    playerID [int] NOT NULL,  
    shirt_number [int] NOT NULL,  
    bench bit,  
    lineup bit,  
    home bit,  
    away bit,
```

```

        event_type    [nvarchar](50),
        event_time    [nvarchar](50),
        player_out_id [int],
        player_out_time    [nvarchar](50)
    )
Else
    Truncate TABLE MatchesPlayers

    Jogadores_Ligas

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'LeaguePlayers')

CREATE TABLE [dbo].[LeaguePlayers](
    player_id    [int] NOT NULL,
    competition_id [int] NOT NULL,
    starting_year [int] NOT NULL,
    ending_year  [int] NOT NULL,
    club_team_id [int] NOT NULL,
    club_team_2_id [int],
    minutes_played_overall    [int] NOT NULL,
    appearances_overall    [int] NOT NULL,
    goals_overall [int] NOT NULL,
    clean_sheets_overall [int] NOT NULL,
    conceded_overall    [int] NOT NULL,
    assists_overall    [int] NOT NULL,
    penalty_goals [int] NOT NULL,
    penalty_misses [int] NOT NULL,
    penalty_success    [int] NOT NULL,
    goals_involved_per_90_overall    [decimal](6,3) NOT NULL,
    assists_per_90_overall    [decimal](6,3) NOT NULL,
    goals_per_90_overall [decimal](6,3) NOT NULL,
    clean_sheets_per_overall    [int] NOT NULL,

```

```

min_per_goal_overall [int] NOT NULL,
conceded_per_90_overall [decimal](6,3) NOT NULL,
min_per_conceded_overall [int] NOT NULL,
cards_overall [int] NOT NULL,
yellow_cards_overall [int] NOT NULL,
red_cards_overall [int] NOT NULL,
min_per_match [int] NOT NULL,
min_per_card_overall [int] NOT NULL,
min_per_assist_overall [int] NOT NULL,
cards_per_90_overall [decimal](6,3) NOT NULL,
rank_in_league_top_attackers [int] NOT NULL,
rank_in_league_top_midfielders [int] NOT NULL,
rank_in_league_top_defenders [int] NOT NULL,
rank_in_club_top_scorer [int] NOT NULL,
last_match_timestamp [int] NOT NULL
)
Else
Truncate TABLE LeaguePlayers

Equipas_Ligas

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'LeagueTeams')
CREATE TABLE [dbo].[LeagueTeams](
team_id[int] NOT NULL,
table_position[int] NOT NULL,
competition_id[int] NOT NULL,
season [nvarchar](50) NOT NULL,
seasonConceded_overall [int] NOT NULL,
seasonGoalsTotal_overall [int] NOT NULL,
seasonScoredNum_overall [int] NOT NULL,
seasonConcededNum_overall [int] NOT NULL,
seasonGoalsMin_overall [int] NOT NULL,

```

seasonScoredMin_overall [int] NOT NULL,
seasonConcededMin_overall [int] NOT NULL,
seasonGoalDifference_overall [int] NOT NULL,
seasonWinsNum_overall [int] NOT NULL,
seasonDrawsNum_overall [int] NOT NULL,
seasonLossesNum_overall [int] NOT NULL,
seasonCS_overall [int] NOT NULL,
seasonCSPercentage_overall [int] NOT NULL,
seasonAVG_overall [decimal](6,3) NOT NULL,
seasonScoredAVG_overall [decimal](6,3) NOT NULL,
seasonConcededAVG_overall [decimal](6,3) NOT NULL,
winPercentage_overall [int] NOT NULL,
drawPercentage_overall [int] NOT NULL,
losePercentage_overall [int] NOT NULL,
cornersAVG_overall [decimal](6,3) NOT NULL,
shotsAVG_overall [decimal](6,3) NOT NULL,
shotsOnTargetAVG_overall [decimal](6,3) NOT NULL,
shotsOffTargetAVG_overall [decimal](6,3) NOT NULL,
possessionAVG_overall [int] NOT NULL,
dangerous_attacks_num_overall [int] NOT NULL,
attacks_num_overall [int] NOT NULL,
dangerous_attacks_avg_overall [decimal](6,3) NOT NULL,
attacks_avg_overall [decimal](6,3) NOT NULL,
cards_for_overall [int],
cards_for_avg_overall [decimal](6,3),
penalties_won_overall [int],
penalties_scored_overall [int],
penalties_missed_overall [int],
shots_per_goals_scored_overall [decimal](6,3),
shots_on_target_per_goals_scored_overall [decimal](6,3),

```

        shot_conversion_rate_overall [decimal](6,3)
    )
    Else

```

```

        Truncate TABLE LeagueTeams
    
```

Anexo 5 – *Scripts* das tabelas do Data Warehouse

Esta secção exhibe as *scripts* SQL utilizadas na criação do armazém de dados da solução desenvolvida descrito na secção 4.4.

DimCountry

```

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimCountry')
BEGIN
    CREATE TABLE [dbo].[DimCountry](
        countryKey [int] IDENTITY(1,1) NOT NULL,
        continent_s [nvarchar](10) NOT NULL,
        continent_name [nvarchar](50) NOT NULL,
        country_ID [int] NOT NULL,
        iso [nvarchar](80) NOT NULL,
        iso_number int,
        country [nvarchar](120) NOT NULL,
        name_pt [nvarchar](120) NOT NULL,
        CONSTRAINT [PK_DimCountry] PRIMARY KEY CLUSTERED
    (
        countryKey ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    )
CREATE NONCLUSTERED INDEX [NonClusteredIndex-countryKey] ON [dbo].[DimCountry]
(
    countryKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON,
OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)

```

END

DimDate

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimDate')

BEGIN

CREATE TABLE [dbo].[DimDate](

[dateKey] [int] IDENTITY(1,1) NOT NULL,

[fullDate] [date] NOT NULL,

[year] [int] NOT NULL,

[semester] [tinyint] NOT NULL,

[trimester] [tinyint] NOT NULL,

[quarter] [tinyint] NOT NULL,

[month] [tinyint] NOT NULL,

[monthName] [nvarchar](10) NOT NULL,

[week] [tinyint] NOT NULL,

[dayNumberOfMonth] [int] NOT NULL,

[dayOfWeek] [nvarchar](20) NOT NULL,

CONSTRAINT [PK_DimDate] PRIMARY KEY CLUSTERED

(

dateKey ASC

)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]

)

CREATE NONCLUSTERED INDEX [NonClusteredIndex-dateKey] ON [dbo].[DimDate]

(

dateKey ASC

)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)

END

DimEvent_Match

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimEvent_Match')
BEGIN
    CREATE TABLE [dbo].[DimEvent_Match](
        eventKey [int] IDENTITY(1,1) NOT NULL,
        event_type [nvarchar](50),
        bench [nvarchar](3) NOT NULL,
        lineup [nvarchar](3) NOT NULL,
        home [nvarchar](3) NOT NULL,
        away [nvarchar](3) NOT NULL

        CONSTRAINT [PK_DimEvent_Match] PRIMARY KEY CLUSTERED
    (
        eventKey ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    )
CREATE NONCLUSTERED INDEX [NonClusteredIndex-eventKey] ON
[dbo].[DimEvent_Match]
(
    eventKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON,
OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)
END
```

DimLeague

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimLeague')
BEGIN
    CREATE TABLE [dbo].[DimLeague](
        competitionKey [int] IDENTITY(1,1) NOT NULL,
        id [int] NOT NULL,
        name [nvarchar](40) NOT NULL,
```

```

        english_name [nvarchar](40) NOT NULL,
        domestic_scale [int] NOT NULL,
        international_scale [int] NOT NULL,
        status [nvarchar](50) NOT NULL,
        startKey [int] NOT NULL,
        endKey [int] NOT NULL,
        image [nvarchar](260) NOT NULL,
        clubNum [int] NOT NULL,
        season [nvarchar](36) NOT NULL,
        effectiveDate date,
        expiredDate date,
        isCurrent [nvarchar](3) NOT NULL,
CONSTRAINT [PK_DimLeague] PRIMARY KEY CLUSTERED
(
    competitionKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
)
CREATE NONCLUSTERED INDEX [NonClusteredIndex-competitionKey] ON
[dbo].[DimLeague]
(
    competitionKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON,
OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)
END

```

DimPlayer

```

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimPlayer')
BEGIN
    CREATE TABLE [dbo].[DimPlayer](
        playerKey [int] IDENTITY(1,1) NOT NULL,

```

```

        id      [int] NOT NULL,
        full_name  [nvarchar](160) NOT NULL,
        first_name [nvarchar](80) NOT NULL,
        last_name  [nvarchar](120) NOT NULL,
        known_as   [nvarchar](140) NOT NULL,
        height [int],
        weight [int],
        position   [nvarchar](40) NOT NULL,
        age      [int],
        countryKey [int] NOT NULL,
        birthdayKey [int] NOT NULL,
        effectiveDate date,
        expiredDate date,
        isCurrent [nvarchar](3) NOT NULL,
CONSTRAINT [PK_DimPlayer] PRIMARY KEY CLUSTERED
(
    playerKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
)
CREATE NONCLUSTERED INDEX [NonClusteredIndex-playerKey] ON [dbo].[DimPlayer]
(
    playerKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON,
OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)
END

DimStadium

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimStadium')
BEGIN
    CREATE TABLE [dbo].[DimStadium](

```

```

        stadiumKey [int] IDENTITY(1,1) NOT NULL,
        stadium_address [nvarchar](110) NOT NULL,
        stadium_name [nvarchar](60) NOT NULL,
        effectiveDate date,
        expiredDate date,
        isCurrent [nvarchar](3) NOT NULL,
CONSTRAINT [PK_DimStadium] PRIMARY KEY CLUSTERED
(
        stadiumKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
)
CREATE NONCLUSTERED INDEX [NonClusteredIndex-stadiumKey] ON [dbo].[DimStadium]
(
        stadiumKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS =
ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)
END

```

DimYear

```

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimYear')
BEGIN
        CREATE TABLE [dbo].[DimYear](
                yearKey [int] IDENTITY(1,1) NOT NULL,
                year [int] NOT NULL,
CONSTRAINT [PK_DimYear] PRIMARY KEY CLUSTERED
(
                yearKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
)

```

```

CREATE NONCLUSTERED INDEX [NonClusteredIndex-yearKey] ON [dbo].[DimYear]
(
    yearKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS =
ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)

END

DimTime_Match

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimTime_Match')
BEGIN
    CREATE TABLE [dbo].[DimTime_Match](
        event_timeKey [int] IDENTITY(1,1) NOT NULL,
        time [int] NOT NULL,
        time_15min [int] NOT NULL,
        time_30min [int] NOT NULL,
        part [int] NOT NULL,
        CONSTRAINT [PK_DimTime_Match] PRIMARY KEY CLUSTERED
    (
        event_timeKey ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    )
    CREATE NONCLUSTERED INDEX [NonClusteredIndex-event_timeKey] ON [dbo].[
DimTime_Match]
(
    event_timeKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON,
OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)

END

DimTeam

```

```

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimTeam')
BEGIN
    CREATE TABLE [dbo].[DimTeam](
        teamKey [int] IDENTITY(1,1) NOT NULL,
        id      [int] NOT NULL,
        name    [nvarchar](110) NOT NULL,
        cleanName [nvarchar](100) NOT NULL,
        age     [int] NOT NULL,
        foundedKey [int] NOT NULL,
        image   [nvarchar](300) NOT NULL,
        CONSTRAINT [PK_DimTeam] PRIMARY KEY CLUSTERED
    (
        teamKey ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    )
CREATE NONCLUSTERED INDEX [NonClusteredIndex-teamKey] ON [dbo].[DimTeam]
(
    teamKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS =
ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)
END

```

DimMatch

```

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'DimMatch')
BEGIN
    CREATE TABLE [dbo].[DimMatch](
        matchKey [int] IDENTITY(1,1) NOT NULL,
        id int NOT NULL,
        match_description nvarchar(350) NOT NULL,

```

```

CONSTRAINT [PK_DimMatch] PRIMARY KEY CLUSTERED
(
    matchKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
)
CREATE NONCLUSTERED INDEX [NonClusteredIndex-matchKey] ON [dbo].[DimMatch]
(
    matchKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS =
ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF)
END

```

FactLeague_Player

```

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'FactLeague_Player')
CREATE TABLE [dbo].[FactLeague_Player](
    playerKey [int] NOT NULL,
    competitionKey [int] NOT NULL,
    teamKey [int] NOT NULL,
    start_yearKey [int] NOT NULL,
    end_yearKey [int] NOT NULL,
    last_matchKey [int] NOT NULL,
    team_2Key [int],
    minutes_played_overall [int] NOT NULL,
    appearances_overall [int] NOT NULL,
    goals_overall [int] NOT NULL,
    clean_sheets_overall [int] NOT NULL,
    conceded_overall [int] NOT NULL,
    assists_overall [int] NOT NULL,
    penalty_goals [int] NOT NULL,
    penalty_misses [int] NOT NULL,

```

```

penalty_success      [int],
goals_involved_per_90_overall      [decimal](6,3) NOT NULL,
assists_per_90_overall      [decimal](6,3) NOT NULL,
goals_per_90_overall [decimal](6,3) NOT NULL,
clean_sheets_per_overall  [int] NOT NULL,
min_per_goal_overall [int] NOT NULL,
conceded_per_90_overall      [decimal](6,3) NOT NULL,
min_per_conceded_overall  [int] NOT NULL,
cards_overall [int] NOT NULL,
yellow_cards_overall [int] NOT NULL,
red_cards_overall  [int] NOT NULL,
min_per_match [int] NOT NULL,
min_per_card_overall [int] NOT NULL,
min_per_assist_overall      [int] NOT NULL,
cards_per_90_overall [decimal](6,3) NOT NULL,
rank_in_league_top_attackers [int],
rank_in_league_top_midfielders      [int],
rank_in_league_top_defenders [int],
rank_in_club_top_scorer      [int],
CONSTRAINT [PK_FactLeague_Player] PRIMARY KEY CLUSTERED
(
    playerKey ASC,
    competitionKey ASC,
    teamKey ASC,
    last_matchKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
)
ELSE
TRUNCATE TABLE FactLeague_Player

```

FactLeague_Team

```
IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'FactLeague_Team')
```

```
CREATE TABLE [dbo].[FactLeague_Team](  
    teamKey[int] NOT NULL,  
    competitionKey[int] NOT NULL,  
    start_yearKey [int] NOT NULL,  
    end_yearKey [int] NOT NULL,  
    table_position[int] NOT NULL,  
    seasonConceded_overall [int] NOT NULL,  
    seasonGoalsTotal_overall [int] NOT NULL,  
    seasonScoredNum_overall [int] NOT NULL,  
    seasonConcededNum_overall [int] NOT NULL,  
    seasonGoalsMin_overall [int] NOT NULL,  
    seasonScoredMin_overall [int] NOT NULL,  
    seasonConcededMin_overall [int] NOT NULL,  
    seasonGoalDifference_overall [int] NOT NULL,  
    seasonWinsNum_overall [int] NOT NULL,  
    seasonDrawsNum_overall [int] NOT NULL,  
    seasonLossesNum_overall [int] NOT NULL,  
    seasonCS_overall [int] NOT NULL,  
    seasonCSPercentage_overall [int] NOT NULL,  
    seasonAVG_overall [decimal](6,3) NOT NULL,  
    seasonScoredAVG_overall [decimal](6,3) NOT NULL,  
    seasonConcededAVG_overall [decimal](6,3) NOT NULL,  
    winPercentage_overall [int] NOT NULL,  
    drawPercentage_overall [int] NOT NULL,  
    losePercentage_overall [int] NOT NULL,  
    shotsAVG_overall [decimal](6,3) NOT NULL,  
    shotsOnTargetAVG_overall [decimal](6,3) NOT NULL,  
    shotsOffTargetAVG_overall [decimal](6,3) NOT NULL,
```

```

        possessionAVG_overall [int] NOT NULL,
        dangerous_attacks_num_overall      [int] NOT NULL,
        attacks_num_overall  [int] NOT NULL,
        dangerous_attacks_avg_overall      [decimal](6,3) NOT NULL,
        attacks_avg_overall  [decimal](6,3) NOT NULL,
        cards_for_overall    [int] NOT NULL,
        cards_for_avg_overall [decimal](6,3) NOT NULL,
        penalties_won_overall [int] NOT NULL,
        penalties_scored_overall  [int] NOT NULL,
        penalties_missed_overall  [int] NOT NULL,
        shots_per_goals_scored_overall    [decimal](6,3),
        shots_on_target_per_goals_scored_overall  [decimal](6,3),
        shot_conversion_rate_overall [decimal](6,3),
CONSTRAINT [PK_FactLeague_Team] PRIMARY KEY CLUSTERED
(
        teamKey ASC,
        competitionKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
)
ELSE
        TRUNCATE TABLE FactLeague_Team

                FactMatch

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'FactMatch')
        CREATE TABLE [dbo].[FactMatch](
                matchKey [int] IDENTITY(1,1) NOT NULL,
                homeKey [int] NOT NULL,
                awayKey [int] NOT NULL,
                stadiumKey [int] NOT NULL,
                competitionKey [int] NOT NULL,

```

dateKey[int] NOT NULL,
id [int] NOT NULL,
homeGoalCount [int] NOT NULL,
awayGoalCount [int] NOT NULL,
totalGoalCount[int] NOT NULL,
team_a_yellow_cards [int] NOT NULL,
team_b_yellow_cards [int] NOT NULL,
team_a_red_cards [int] NOT NULL,
team_b_red_cards [int] NOT NULL,
team_a_shotsOnTarget [int],
team_b_shotsOnTarget [int],
team_a_shotsOffTarget [int],
team_b_shotsOffTarget [int],
team_a_shots [int],
team_b_shots [int],
team_a_fouls [int],
team_b_fouls [int],
team_a_possession [int],
team_b_possession [int],
team_a_cards_num [int] NOT NULL,
team_b_cards_num [int] NOT NULL,
team_a_dangerous_attacks [int] NOT NULL,
team_b_dangerous_attacks [int] NOT NULL,
team_a_attacks [int] NOT NULL,
team_b_attacks [int] NOT NULL,
team_a_penalties_won [int] NOT NULL,
team_b_penalties_won [int] NOT NULL,
team_a_penalty_goals [int] NOT NULL,
team_b_penalty_goals [int] NOT NULL,
team_a_penalty_missed [int] NOT NULL,

```

        team_b_penalty_missed [int] NOT NULL,
CONSTRAINT [PK_FactMatch] PRIMARY KEY CLUSTERED
(
    homeKey ASC,
    awayKey ASC,
    competitionKey ASC,
    dateKey ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
)
ELSE
    TRUNCATE TABLE FactMatch

FactMatch_Player

IF NOT EXISTS (SELECT name FROM sys.tables WHERE name = 'FactMatch_Player')
CREATE TABLE [dbo].[FactMatch_Player](
    playerKey [int] NOT NULL,
    eventKey [int] NOT NULL,
    event_timeKey [int] NOT NULL,
    matchKey [int] NOT NULL,
    competitionKey [int] NOT NULL,
    stadiumKey [int] NOT NULL,
    homeKey [int] NOT NULL,
    awayKey [int] NOT NULL,
    shirt_number [int],
    player_outKey [int],
    player_out_timeKey [int],
CONSTRAINT [PK_FactMatch_Player] PRIMARY KEY CLUSTERED
(
    playerKey ASC,
    eventKey ASC,

```

```

        event_timeKey ASC,

        matchKey ASC

    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY =
OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]

)

ELSE

    TRUNCATE TABLE FactMatch_Player

```

Anexo 6 – Inquérito de Satisfação sobre a Solução

Este anexo exhibe o questionário de satisfação que serviu para avaliação da solução desenvolvida.

Inquérito de Satisfação sobre a Solução / Solution Satisfaction Survey

B *I* U

Este inquérito tem como objetivo recolher opiniões e sugestões sobre a solução de Business Intelligence implementada, que inclui dashboards com indicadores de desempenho de atletas e equipas em jogos e competições de futebol. De forma a avaliar a solução é utilizada uma escala de 1 a 5, sendo que o valor 1 representa "Discordo Totalmente", 2 "Discordo", 3 "Neutro", 4 "Concordo" e 5 "Concordo Totalmente".

Agradecemos o seu tempo e colaboração. As suas respostas são valiosas para nós e contribuirão significativamente para a melhoria contínua da nossa solução.

This survey aims to gather opinions and suggestions on the Business Intelligence solution implemented, which includes dashboards with performance indicators for athletes and teams in soccer matches and competitions. In order to evaluate the solution, a scale of 1 to 5 is used, with 1 representing "Strongly Disagree", 2 "Disagree", 3 "Neutral", 4 "Agree" and 5 "Strongly Agree".

Thank you for your time and cooperation. Your responses are valuable to us and will contribute significantly to the continuous improvement of our solution.

Qual é o seu papel no clube? (ex.: treinador, analista, treinador adjunto, outro) *
 What is your role at the club? (e.g. coach, analyst, assistant coach, other)

Texto de resposta curta

Licença / License *

- Nível/Level 2 (UEFA B / Conmebol B / AFC B)
- Nível/Level 3 (UEFA A / Conmebol A / AFC A)
- Nível/Level 4 (UEFA Pro / Conmebol Pro / AFC Pro)
- Outra opção...

Qual dispositivo utilizou para consultar o dashboard? *

Which device did you use to consult the dashboard?

- Portátil / Computer
- Telémoveel / Phone
- Tablet
- Ipad
- Outra: _____

Qual o sistema operativo do dispositivo que utilizou para consultar o dashboard? *

Which device operation system did you use to consult the dashboard?

- Windows
- MacOS (Apple)
- Android
- iOS (Apple Phone)

1. O painel (dashboard) é intuitivo e fácil de usar? *

1. Is the dashboard intuitive and easy to use?

- 1
- 2
- 3
- 4
- 5

2. A navegação entre os diferentes dashboards é clara e sem complicações? *

2. Is the navigation between the different dashboards clear and uncomplicated?

- 1
- 2
- 3
- 4
- 5

3. Os indicadores apresentados são facilmente compreensíveis? *

3. Are the indicators shown easy to understand?

- 1
- 2
- 3
- 4
- 5

4. Os dados apresentados nos dashboards são precisos e confiáveis? *

4. Is the data presented in the dashboards accurate and reliable?

- 1
- 2
- 3
- 4
- 5

5. Os dados são atualizados com a frequência necessária? *

5. Is the data updated as often as necessary?

- 1
- 2
- 3
- 4
- 5

6. Considera que os dados históricos apresentados encontram-se integros? *

6. Do you consider the historical data presented to be complete?

- 1
- 2
- 3
- 4
- 5

7. O tempo de resposta ao carregar os dashboards é satisfatório? *

7. Is the response time when loading dashboards satisfactory?

- 1
- 2
- 3
- 4
- 5

8. Não há atrasos perceptíveis na visualização dos dados ao interagir com os dashboards? *

8. Are there no noticeable delays in viewing the data when interacting with the dashboards?

- 1
- 2
- 3
- 4
- 5

9. Os dashboards funcionaram corretamente no dispositivo que utilizou? *

9. Did the dashboards work correctly on the device you used?

- 1
- 2
- 3
- 4
- 5

10. Os dashboards funcionam corretamente no sistema operativo que utilizou? *

10. Do the dashboards work correctly on the operating system you used?

- 1
- 2
- 3
- 4
- 5

11. Os dados históricos disponíveis são suficientes para análises e comparações? *

11. Is the historical data available sufficient for analysis and comparison?

- 1
- 2
- 3
- 4
- 5

12. A apresentação do histórico de dados facilita a compreensão das tendências de desempenho? *

12. Does the presentation of historical data make it easier to understand performance trends?

- 1
- 2
- 3
- 4
- 5

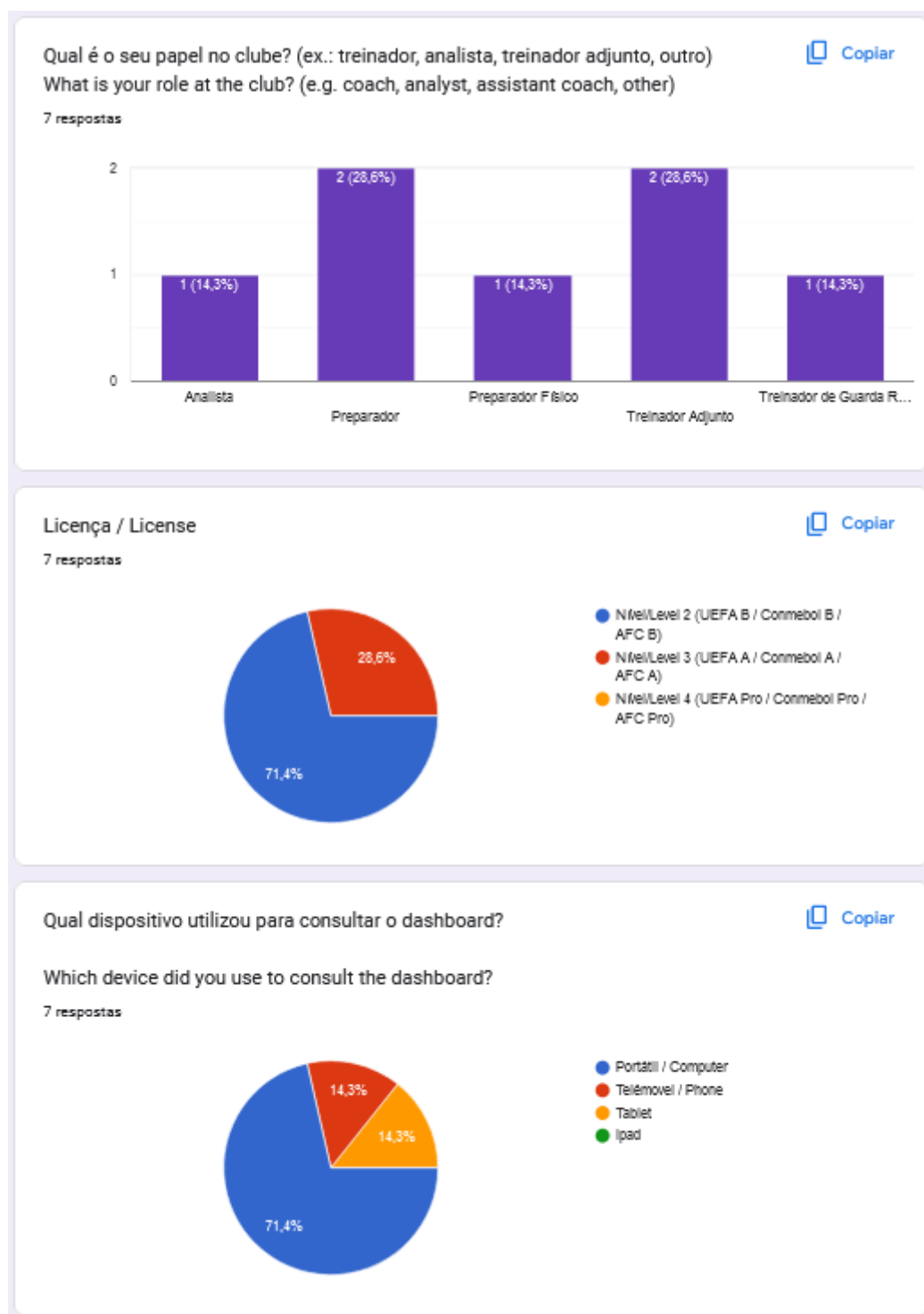
Tem algum comentário ou sugestão para melhorar a solução que lhe foi apresentada? *

Do you have any comments or suggestions for improving the solution presented to you?

A sua resposta

Anexo 7 – Resultados do inquérito de satisfação

Este anexo exhibe os resultados do questionário de satisfação presente no Anexo 6.

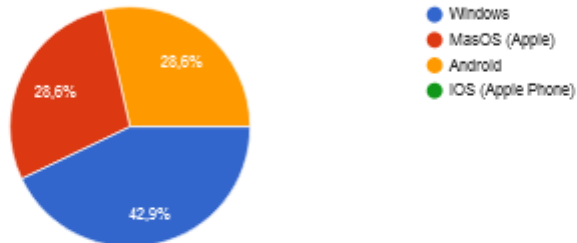


Qual o sistema operativo do dispositivo que utilizou para consultar o dashboard?

[Copiar](#)

Which device operation system did you use to consult the dashboard?

7 respostas

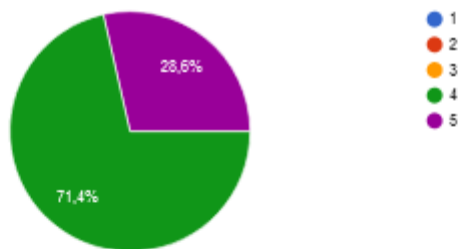


1. O painel (dashboard) é intuitivo e fácil de usar?

[Copiar](#)

1. Is the dashboard intuitive and easy to use?

7 respostas

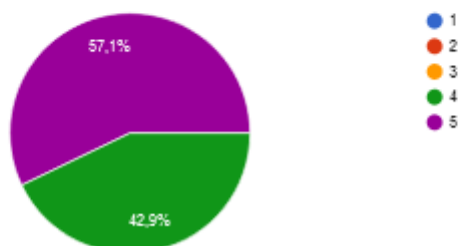


2. A navegação entre os diferentes dashboards é clara e sem complicações?

[Copiar](#)

2. Is the navigation between the different dashboards clear and uncomplicated?

7 respostas

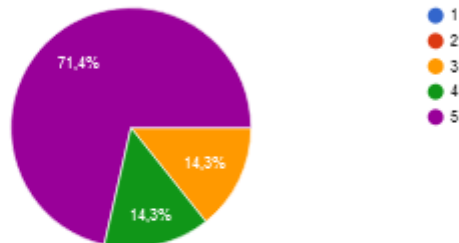


3. Os indicadores apresentados são facilmente compreensíveis?

[Copiar](#)

3. Are the indicators shown easy to understand?

7 respostas

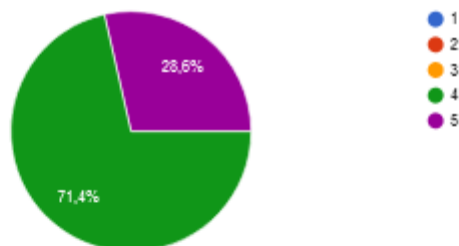


4. Os dados apresentados nos dashboards são precisos e confiáveis?

[Copiar](#)

4. Is the data presented in the dashboards accurate and reliable?

7 respostas

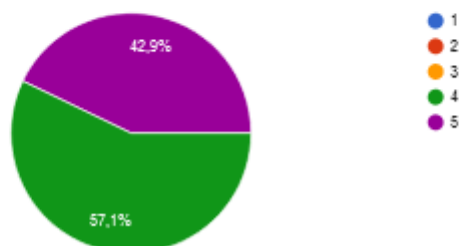


5. Os dados são atualizados com a frequência necessária?

[Copiar](#)

5. Is the data updated as often as necessary?

7 respostas

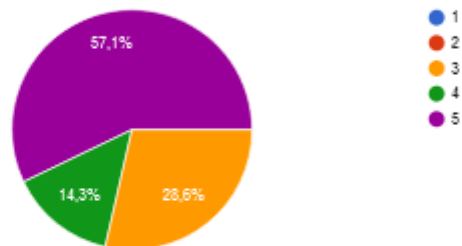


6. Considera que os dados históricos apresentados encontram-se integros?

[Copiar](#)

6. Do you consider the historical data presented to be complete?

7 respostas

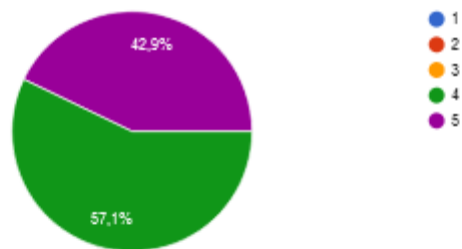


7. O tempo de resposta ao carregar os dashboards é satisfatório?

[Copiar](#)

7. Is the response time when loading dashboards satisfactory?

7 respostas

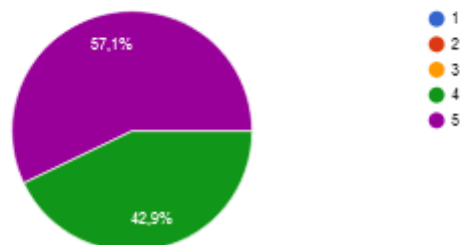


8. Não há atrasos perceptíveis na visualização dos dados ao interagir com os dashboards?

[Copiar](#)

8. Are there no noticeable delays in viewing the data when interacting with the dashboards?

7 respostas

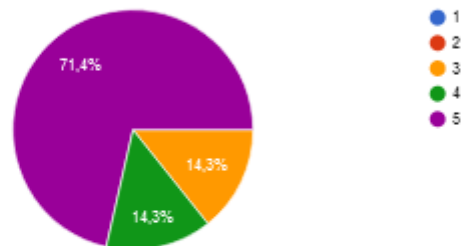


9. Os dashboards funcionaram corretamente no dispositivo que utilizou?

[Copiar](#)

9. Did the dashboards work correctly on the device you used?

7 respostas

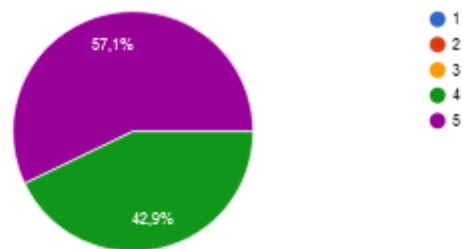


10. Os dashboards funcionam corretamente no sistema operativo que utilizou?

[Copiar](#)

10. Do the dashboards work correctly on the operating system you used?

7 respostas

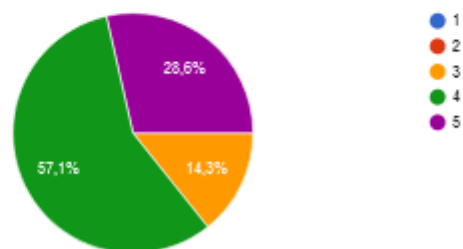


11. Os dados históricos disponíveis são suficientes para análises e comparações?

[Copiar](#)

11. Is the historical data available sufficient for analysis and comparison?

7 respostas

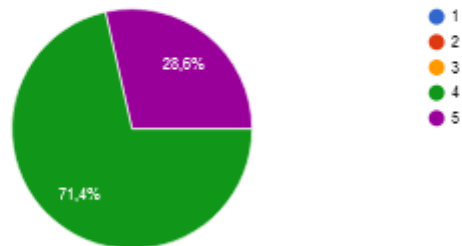


12. A apresentação do histórico de dados facilita a compreensão das tendências de desempenho?

 Copiar

12. Does the presentation of historical data make it easier to understand performance trends?

7 respostas



Tem algum comentário ou sugestão para melhorar a solução que lhe foi apresentada?

Do you have any comments or suggestions for improving the solution presented to you?

7 respostas

Deveria incluir-se outros dados, como passes, passes errados, passes certos, distancia percorrida e coordenadas dentro de uma partida.

Poderiam criar mais dashboards para grupo de posição (guardar-redes, Defesa, Médios, Atacantes) com dados distintos

Deveria ser possível verificar os dados ao nível mensal, para percebermos a tendencia mensal. Para além disso, uma pessoa que não saiba onde a liga que uma equipa ou varias equipas jogam fica dificil selecionar. Deveria ser possível filtrar as equipas através da competição selecionada, e com isso os adversários também.

Nem todos os dados apresentam o mesmo volume de histórico

Filtro de Equipas por Ligas/Competição

Diminuição de quantidades dos anos possíveis de escolha, uma vez que não apresentam dados.

Escolha da Liga filtrar as equipas a selecionar