

COIMBRA
BUSINESS
SCHOOL

 **iscac** 
Politécnico de Coimbra

**COIMBRA
BUSINESS
SCHOOL**
 **iscac** 
Politécnico de Coimbra

Ana Beatriz Oliveira Gaspar

**Desenvolvimento de Modelos Semânticos:
*Business Intelligence na Zeone Consulting***

Coimbra, outubro de 2024



Ana Beatriz Oliveira Gaspar

Desenvolvimento de Modelos Semânticos: *Business Intelligence na Zeone Consulting*

Relatório de estágio submetido ao Instituto Superior de Contabilidade e Administração de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de **Mestre em Análise de Dados e Sistemas de Apoio à Decisão**, realizado sob a orientação da Professora Doutora Joana Jorge de Queiroz Leite e da Professora Doutora Dora Regina Oliveira Melo e supervisão de Dr. António Manuel de Matos Rodrigues.

Coimbra, outubro de 2024

TERMO DE RESPONSABILIDADE

Declaro ser a autora deste relatório de estágio, que constitui um trabalho original e inédito, que nunca foi submetido a outra Instituição de ensino superior para obtenção de um grau académico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas e que tenho consciência de que o plágio constitui uma grave falta de ética, que poderá resultar na anulação do presente relatório de estágio.

AGRADECIMENTOS

Concluir esta etapa é o resultado de um longo percurso que não teria sido possível sem o apoio e a presença de várias pessoas, às quais estou profundamente grata.

Em primeiro lugar, ao meu pai e aos meus avós, pelas suas presenças incondicionais e por me terem apoiado nos momentos em que mais precisei. As suas forças foram fundamentais para que eu nunca desistisse dos meus objetivos.

Ao meu namorado Hugo, pelo amor, carinho, apoio incansável, pela paciência inesgotável e por me motivar diariamente. Obrigada por todos os "hoje fazes mais um bocadinho", que, somados, me trouxeram até aqui.

Às minhas orientadoras, a Professora Doutora Joana Jorge de Queiroz Leite e a Professora Doutora Dora Regina Oliveira Melo, pela disponibilidade constante, compreensão e paciência com os meus prazos. A vossa orientação foi essencial para que este trabalho tomasse forma.

Ao meu supervisor Dr. António Manuel de Matos Rodrigues, por me acolher e me guiar com sabedoria, e ao meu colega André Santos, por estar sempre disponível para esclarecer dúvidas, proporcionando-me um ambiente de aprendizagem e crescimento.

Aos meus colegas da *Zeone*, que me receberam de braços abertos e me fizeram sentir parte de uma verdadeira família, o meu mais sincero agradecimento.

À Catarina Melo e ao João Pinto, meus colegas de estágio, por todos os momentos partilhados, desde o primeiro dia até ao último. A vossa companhia fez desta jornada uma experiência muito mais leve e especial.

Por fim, a todos aqueles que me acompanharam ao longo deste mestrado, um enorme obrigada. Cada lição, cada gesto, e cada palavra de encorajamento contribuiu para o que sou hoje.

RESUMO

Os Sistemas de Apoio à Decisão, alicerçados em *Business Intelligence*, são essenciais para que as organizações possam integrar, visualizar e explorar dados de diferentes fontes para decisões rápidas e informadas. O *Power BI* destaca-se como um *software* de *Business Intelligence* eficaz na criação de relatórios que facilitam o acesso a indicadores relevantes. Este *software* desempenha um papel central no desenvolvimento de modelos semânticos que permitem estruturar e relacionar dados, assegurando uma análise consistente e acessível aos utilizadores. A criação de modelos semânticos no *Power BI* é crucial para garantir a integridade e a qualidade dos dados utilizados, facilitando uma interpretação precisa e fiável das informações. Além disso, a aplicação de boas práticas no desenvolvimento de relatórios, como a organização lógica dos dados e o uso de *storytelling*, torna as visualizações mais intuitivas e relevantes para o processo de tomada de decisão. Desta forma, o *Power BI*, ao integrar qualidade de dados e técnicas eficazes de apresentação, torna-se uma ferramenta essencial para suportar decisões informadas e estratégicas.

É neste contexto que se insere o estágio realizado na *Zeone Consulting*, empresa do Grupo FHC. Os trabalhos realizados focaram-se no desenvolvimento de modelos semânticos e envolveram a preparação e integração de dados de várias fontes, a criação de modelos entidade-relacionamento e o desenvolvimento de medidas personalizadas em DAX para gerar relatórios interativos. Um dos principais desafios foi garantir a qualidade dos dados utilizados nas métricas, o que exigiu validação constante. A criação de relatórios claros e interativos, com o uso de boas práticas de *storytelling*, foi fundamental para apoiar a tomada de decisões informadas. As conclusões destacam a importância da qualidade dos dados, e sugerem-se futuros desenvolvimentos como a automação de processos e a integração de técnicas de inteligência artificial para otimizar as análises.

Palavras-chave:

Sistemas de Apoio à Decisão, *Business Intelligence*, *Power BI*, Modelos Semânticos, Qualidade dos Dados

ABSTRACT

Decision Support Systems, supported by Business Intelligence, are essential for organizations to integrate, visualize, and explore data from various sources, enabling quick and informed decisions. Power BI stands out as an effective Business Intelligence software for creating reports that facilitate access to relevant indicators. This software plays a central role in developing semantic models that structure and relate data, ensuring consistent and accessible analysis for users. Creating semantic models in Power BI is crucial to maintaining the integrity and quality of data, allowing for accurate and reliable information interpretation. Additionally, applying best practices in report development, such as logical data organization and storytelling, makes visualizations more intuitive and relevant to the decision-making process. In this way, Power BI, by integrating data quality and effective presentation techniques, becomes an essential tool for supporting informed and strategic decisions.

This approach aligns with the internship carried out at Zeone Consulting, a company within the FHC Group. The work focused on developing semantic models and involved data preparation and integration from various sources, creating entity relationship models, and developing custom DAX measures to generate interactive reports. One of the main challenges was ensuring the quality of the data used in metrics, requiring constant validation. Creating clear and interactive reports using storytelling best practices was fundamental to supporting informed decision-making. The conclusions highlight the importance of data quality and suggest future developments, such as process automation and the integration of artificial intelligence techniques to optimize analyses.

Keywords:

Decision Support Systems, Business Intelligence, Power BI, Semantic Models, Data Quality

ÍNDICE GERAL

| | |
|-----------------------------------------------------------------------------------------|----|
| INTRODUÇÃO..... | 1 |
| 1 Os Sistemas de Apoio à Decisão e as Ferramentas de <i>Business Intelligence</i> | 3 |
| 1.1 Contextualização..... | 3 |
| 1.2 Tomada de Decisão na Gestão de Negócio e Sistemas de Apoio à Decisão | 3 |
| 1.3 <i>Business Intelligence</i> e as capacidades analíticas | 5 |
| 1.3.1 Integração de BI com Tecnologias Emergentes | 6 |
| 1.3.2 Melhores Práticas para uma Implementação Eficaz de BI..... | 7 |
| 1.4 Qualidade dos dados | 8 |
| 1.5 Boas práticas de desenvolvimento de relatórios e <i>storytelling</i> | 11 |
| 1.6 <i>Power BI</i> e modelos semânticos | 14 |
| 2 <i>Zeone Consulting</i> | 19 |
| 2.1 Apresentação..... | 19 |
| 2.2 Estágio | 21 |
| 3 Metodologia do trabalho..... | 23 |
| 3.1 Metodologia CRISP-DM | 23 |
| 3.2 Metodologia aplicada ao contexto do estágio..... | 24 |
| 4 Desenvolvimento de Modelos Semânticos..... | 27 |
| 4.1 Enquadramento das atividades desenvolvidas nas fases do CRISP-DM..... | 27 |
| 4.2 Análise e entendimento do negócio | 28 |
| 4.3 Seleção e entendimento dos dados..... | 29 |
| 4.4 Preparação dos dados | 30 |
| 4.5 Modelação..... | 36 |

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na *Zeone Consulting*

| | | |
|----------------------------------|-------------------------------------|----|
| 4.5.1 | Modelos ER | 36 |
| 4.5.2 | Medidas em DAX..... | 38 |
| 4.5.3 | Desenvolvimento de relatórios | 58 |
| 4.6 | Avaliação dos modelos | 65 |
| 4.7 | Implantação..... | 67 |
| CONCLUSÃO..... | | 71 |
| REFERÊNCIAS BIBLIOGRÁFICAS | | 72 |

ÍNDICE DE FIGURAS

| | |
|----------------------------------------------------------------------------------------|----|
| Figura 2.1- Posicionamento da Zeone Consulting no Grupo FHC..... | 20 |
| Figura 3.1 - Fases do ciclo de vida da metodologia CRISP-DM. | 23 |
| Figura 4.1- Tabela de datas e parâmetros correspondentes | 30 |
| Figura 4.2- Coluna "DataAdmissao" no formato original..... | 31 |
| Figura 4.3- Coluna "DataAdmissao" após alteração do tipo de data..... | 31 |
| Figura 4.4- Criação de coluna calculada em Power Query | 31 |
| Figura 4.5- Coluna "_Idade" como resultado da criação de coluna calculada | 32 |
| Figura 4.6- Coluna "_Idade" após alteração de tipo de dados..... | 32 |
| Figura 4.7- Tabela "ZZUtilitieRubrica" após importação em Power Query | 33 |
| Figura 4.8- Tabela "ZZUtilitieRubrica" após remoção de colunas desnecessárias | 33 |
| Figura 4.9- Coluna "zeo_artigos(zeo_artigos)" após importação em Power Query..... | 34 |
| Figura 4.10- Coluna "zeo_artigos(zeo_artigos)" expandida..... | 34 |
| Figura 4.11- Coluna "PlantName - Copy" com valores extraídos entre delimitadores | 35 |
| Figura 4.12- Coluna "Unit" com alteração de valores..... | 35 |
| Figura 4.13- Criação de coluna calculada condicional "Património Genético" | 36 |
| Figura 4.14- Modelo ER: Star Schema..... | 37 |
| Figura 4.15- Modelo ER: Snowflake Schema | 38 |
| Figura 4.16- Medida "Total Salary Cost" | 39 |
| Figura 4.17- Medida "Total Consumos N" | 39 |
| Figura 4.18- Medida "Total Consumos N-1" | 40 |
| Figura 4.19- Medida "Varia Indicadores" | 41 |
| Figura 4.20- Medida "Headcount" | 42 |

| | |
|-----------------------------------------------------------------------|----|
| Figura 4.21- Medida "Intensidade Energética 20" | 42 |
| Figura 4.22- Power BI: Eixo Y Constante..... | 43 |
| Figura 4.23- Medida "Total Salary Cost N-1" | 43 |
| Figura 4.24- Medida "YTD Overtime (hrs)" | 44 |
| Figura 4.25- Medida "eixo intensidade energética" | 44 |
| Figura 4.26- Medida "eixo valorização"..... | 45 |
| Figura 4.27- Medida "Consumo Ano Title" | 45 |
| Figura 4.28- Resultado do título após selecionar "TEPs" | 46 |
| Figura 4.29- Resultado do título após selecionar "SGCIE"..... | 47 |
| Figura 4.30- Resultado do título após selecionar "INERPA & CELE"..... | 47 |
| Figura 4.31- Medida "CF Amarelo" | 48 |
| Figura 4.32- Medida "CF Verde" | 48 |
| Figura 4.33- Medida "CF Vermelho" | 48 |
| Figura 4.34- Medida "Color Varia Indicadores" | 48 |
| Figura 4.35- Medida "Color waste varia n-2/n-1" | 49 |
| Figura 4.36- Medida "Valorizacao"..... | 50 |
| Figura 4.37- Medida "m3 normalizado"..... | 51 |
| Figura 4.38- Medida "Top N ECHA"..... | 52 |
| Figura 4.39- Resultado do gráfico quando selecionado o Top 5 | 52 |
| Figura 4.40- Resultado do gráfico quando selecionado Top 8..... | 53 |
| Figura 4.41- Tabela calculada "zHeadcount" - Parte 1..... | 53 |
| Figura 4.42- Tabela calculada "zHeadcount" - Parte 2..... | 54 |
| Figura 4.43- Medida "Base salary (nominal)" | 55 |
| Figura 4.44- Medida "Overtime (hrs)" | 56 |

| | |
|--------------------------------------------------------------------------------------------|----|
| Figura 4.45- Medida "SKU from several regions" | 57 |
| Figura 4.46- Dashboard do relatório "People Cost" | 58 |
| Figura 4.47- Analytics do relatório "People Cost" | 59 |
| Figura 4.48- Validation do relatório "People Cost" | 60 |
| Figura 4.49- Detailed do relatório "People Cost" | 61 |
| Figura 4.50- Dashboard com Slicers do relatório "Fluorinated Gases" | 62 |
| Figura 4.51- KPI Financeiros com Tooltip no relatório "Strategic Planning" | 62 |
| Figura 4.52- Painel de filtros aplicados diretamente na página..... | 63 |
| Figura 4.53- Valores em branco no relatório "People Cost" | 63 |
| Figura 4.54- Formatação de cores e Tooltip com fórmulas no relatório "Utilities" | 64 |
| Figura 4.55- Formatação de cor das células da tabela no relatório "Fluorinated Gases" | 64 |
| Figura 4.56- Medida de validação de número artigos no relatório "Chemical Products" ... | 65 |
| Figura 4.57- Medida de validação após correção no relatório "Chemical Products" | 66 |
| Figura 4.58- Tabela dinâmica para validação da coluna "_Quantidade KILO" | 66 |
| Figura 4.59- Verificação do resultado da medida criada..... | 67 |
| Figura 4.60 - Security Roles | 68 |
| Figura 4.61- Verificação de documentos de valorização em falta | 69 |
| Figura 4.62- Introdução de dados incorretos no campo "Utility Forn." | 69 |
| Figura 4.63- Verificação de documentos após atualização dos dados | 70 |

ÍNDICE DE TABELAS

| | |
|--------------------------------------------------------------------------------|----|
| Tabela 3.1 - Fases adaptadas do CRISP-DM usadas em cada modelo semântico. | 27 |
|--------------------------------------------------------------------------------|----|

Lista de abreviaturas, acrónimos e siglas

| | |
|----------|--------------------------------------------------------|
| BI | <i>Business Intelligence</i> |
| CRISP-DM | <i>CRoss Industry Standard Process for Data Mining</i> |
| DAX | <i>Data Analysis Expressions</i> |
| DM | <i>Data Mining</i> |
| ER | Entidade-Relacionamento |
| ETL | <i>Extract, Transform, Load</i> |
| KPI | <i>Key Performance Indicator</i> |
| RLS | <i>Row Level Security</i> |
| SQL | <i>Structured Query Language</i> |

INTRODUÇÃO

A análise de dados é uma componente muito importante no processo de transformação digital nas empresas, sendo fundamental para otimizar processos, melhorar a tomada de decisões e fornecer indicadores valiosos (Romero et al., 2019). O uso de ferramentas de Inteligência de Negócio (BI, do inglês *Business Intelligence*), como o *software Power BI*, tem-se mostrado essencial para que as organizações possam integrar, visualizar e explorar dados de diferentes fontes. No contexto empresarial, a garantia da qualidade e integridade dos dados, bem como a implementação de boas práticas de análise, são essenciais para assegurar que as decisões sejam informadas e baseadas em dados fiáveis e precisos. As práticas de BI são cada vez mais necessárias para que as empresas possam integrar, visualizar e explorar dados de diferentes fontes, possibilitando uma análise mais profunda e uma tomada de decisão mais assertiva (Laudon & Laudon, 2022).

A *Zeone Consulting*, empresa do Grupo FHC, em plena transição digital, abriu a oportunidade da realização de um estágio com o propósito de desenvolver e implementar soluções analíticas que assegurassem a integração de boas práticas de BI, promovessem a qualidade e integridade dos dados e permitissem uma tomada de decisão mais eficiente.

Neste enquadramento, o principal objetivo definido para este estágio foi o desenvolvimento de soluções analíticas no *Power BI* que permitissem a criação de modelos semânticos bem estruturados e visualizações interativas. Especificamente, o estágio focou-se no:

- Desenvolvimento de modelos semânticos através da criação de modelos Entidade-Relacionamento (ER) e medidas em Expressões de Análise de Dados (DAX, do inglês *Data Analysis Expressions*) para facilitar a análise e interpretação dos dados;
- Garantia da qualidade dos dados, através de processos de validação e *storytelling* visual para comunicar resultados de forma clara.

A metodologia adotada para a concretização do objetivo estabelecido foi a metodologia *CRoss Industry Standard Process for Data Mining* (CRISP-DM), amplamente utilizada para estruturar projetos de análise de dados na Indústria. Esta metodologia é composta por seis fases (Chapman et al., 2000): compreensão do negócio, compreensão dos dados,

preparação dos dados, modelação, avaliação e implantação. Embora originalmente concebido para projetos de mineração de dados (DM, do inglês *data mining*), o CRISP-DM foi adaptado para o trabalho realizado em *Power BI*, onde o foco esteve na criação de modelos semânticos com relatórios interativos e na garantia da qualidade dos dados, em vez do DM propriamente dita.

No que se segue, o relatório está organizado em cinco capítulos. O primeiro capítulo aborda os conceitos teóricos fundamentais à análise de dados e às práticas de BI, com base em referências bibliográficas relevantes. São descritas ferramentas de análise de dados, como o *software Power BI*, e a importância da qualidade dos dados, bem como as melhores práticas para o desenvolvimento de relatórios interativos.

O segundo capítulo apresenta uma visão geral da *Zeone Consulting* e do Grupo FHC, contextualizando o estágio e os processos internos da empresa. Este capítulo foca-se na transição digital em curso na empresa, na integração de dados de diversas fontes e na necessidade de garantir a sua integridade e confidencialidade.

O terceiro capítulo descreve como a metodologia CRISP-DM foi adaptada ao trabalho realizado no *Power BI*, explicando de que forma as suas fases foram aplicadas na construção e desenvolvimento de modelos semânticos e desenvolvimento de relatórios interativos, embora não tenha havido DM no projeto.

O quarto capítulo detalha as atividades desenvolvidas durante o estágio, com um foco particular na aplicação prática das fases do CRISP-DM, como a preparação de dados, modelação e avaliação de modelos, e também nas soluções implementadas para garantir a qualidade dos dados e a usabilidade dos relatórios no contexto empresarial.

Por fim, apresenta-se as conclusões do estágio, identificando os principais desafios enfrentados, como a necessidade de garantir a qualidade dos dados. São também apresentadas as recomendações para trabalhos futuros, sugerindo melhorias nos processos de integração de dados, automação de relatórios, e a exploração de técnicas mais avançadas, como a integração de Inteligência Artificial para otimizar as análises.

1 Os Sistemas de Apoio à Decisão e as Ferramentas de *Business Intelligence*

Este capítulo de enquadramento teórico vai focar-se nas cinco áreas centrais que suportam os objetivos do estágio, concretamente: *Business Intelligence*, *Power BI* e Modelos Semânticos, Qualidade dos Dados e Boas Práticas de Desenvolvimento de Relatórios.

1.1 Contextualização

Os Sistemas de Apoio à Decisão, suportados em BI, desempenham um papel fundamental na gestão empresarial moderna, onde tomadas de decisão rápidas e informadas são essenciais para a competitividade organizacional. A integração de Sistemas de Informação de Gestão com Sistemas de Apoio à Decisão é crucial para melhorar os processos de decisão, proporcionando às organizações os benefícios necessários para enfrentar os desafios do mercado (Laudon & Laudon, 2022).

O impacto das tecnologias de BI e das técnicas de Análise de Negócio na tomada de decisões é significativo para melhorar as capacidades analíticas (Laudon & Laudon, 2022). Neste contexto, o *software Power BI*, a sua utilização como plataforma para o desenvolvimento de visualizações integradas de *dashboards* de desempenho, tem demonstrado ser uma ferramenta eficaz para a criação de relatórios e para a partilha de indicadores relevantes (Dhanoa et al., 2024).

Além disso, a gestão e a qualidade dos dados são importantes para assegurar que as análises realizadas são baseadas em informações objetivas, credíveis e fiáveis (Rubin, 2014). Como forma de boa prática de desenvolvimento de relatórios, a comunicação dos dados através do *storytelling* desempenha um papel essencial, facilitando a interpretação dos indicadores apresentados e melhorando o processo de decisão nas organizações (Dhanoa et al., 2024).

1.2 Tomada de Decisão na Gestão de Negócio e Sistemas de Apoio à Decisão

O processo de tomada de decisão nas organizações tem vindo a sofrer uma transformação significativa ao longo das últimas décadas. No passado, a responsabilidade pelas decisões estava concentrada nas mãos da gestão de topo. Contudo, com o avanço dos sistemas de

informação, esta responsabilidade foi gradualmente descentralizada, abrangendo agora vários níveis hierárquicos da empresa. Estes sistemas disponibilizam informações relevantes a um maior número de utilizadores, permitindo uma abordagem mais inclusiva e eficaz na tomada de decisões. Os gestores desempenham um conjunto de papéis fundamentais nas organizações, que podem ser agrupados em três categorias, de acordo com a tipologia dos autores Laudon e Laudon (2022): interpessoais, informacionais e decisivos. Nos papéis interpessoais, os gestores representam a organização em eventos externos, motivam e orientam colaboradores, e mediam as relações entre diferentes níveis hierárquicos. Nos papéis informacionais, os gestores atuam como centros de informação, recolhendo e distribuindo dados relevantes para a organização. Nos papéis decisivos, os gestores tomam decisões estratégicas, gerem crises, afetam recursos e mediam conflitos.

De acordo com Laudon e Laudon (2022), as decisões dentro das organizações podem ser classificadas em três categorias principais: não estruturadas, estruturadas e semiestruturadas. As decisões não estruturadas são aquelas que exigem uma análise mais subjetiva, baseada em julgamento, avaliação e intuição. Estas decisões são, por norma, únicas, de grande importância e não rotineiras, sendo predominantemente tomadas por gestores de topo. Um exemplo seria a definição de metas estratégicas a longo prazo ou a decisão de entrada em novos mercados. Por outro lado, as decisões estruturadas caracterizam-se por serem repetitivas e rotineiras, seguindo procedimentos claramente definidos. Estas são mais comuns nos níveis operacionais da organização, como a determinação da elegibilidade para horas extraordinárias ou a concessão de crédito a clientes, baseando-se em critérios preestabelecidos. As decisões semiestruturadas combinam elementos das decisões estruturadas e não estruturadas. Um exemplo típico seria o de um gestor intermédio que determina um motivo para uma queda nas vendas, recorrendo a dados estruturados e a informações mais informais, como as tendências económicas locais. Embora o impacto de cada decisão individual possa parecer pequeno, a melhoria contínua em múltiplas pequenas decisões ao longo do tempo pode gerar um valor substancial para a organização.

Laudon e Laudon (2022) referem ainda que o processo de tomada de decisão envolve quatro fases principais: inteligência, *design*, escolha e implantação. A fase de inteligência envolve a identificação e compreensão dos problemas existentes na organização,

analisando as suas causas e os impactos. Na fase de *design*, elaboram-se e avaliam-se possíveis soluções para o problema identificado. A escolha consiste na seleção da melhor alternativa entre as opções disponíveis. A implantação envolve a execução da solução escolhida e a monitorização contínua dos seus resultados. Caso a solução implementada não produza os resultados esperados, pode ser necessário voltar a etapas anteriores do processo para ajustar ou identificar uma nova abordagem.

No entanto, a eficácia deste processo depende da qualidade dos sistemas de informação utilizados para apoiar a tomada de decisão. Embora os sistemas de informação possam melhorar a eficiência e a qualidade das decisões, a sua eficácia não é garantida em todos os contextos. Existem três fatores principais que podem limitar a qualidade das decisões tomadas: a qualidade da informação, as perceções dos gestores e a inércia e política organizacionais. As decisões eficazes requerem informação precisa e oportuna, mas dados imprecisos ou desatualizados podem comprometer a qualidade da decisão. Além disso, os gestores processam informações através de filtros cognitivos que podem distorcer a perceção da realidade, influenciando negativamente as decisões. Por fim, as organizações, muitas vezes, resistem a mudanças, e as decisões podem ser influenciadas por dinâmicas políticas internas, em vez de se focarem nas melhores soluções para os problemas identificados (Laudon & Laudon, 2022).

No contexto atual, muitas decisões organizacionais são automatizadas e processadas por algoritmos, especialmente em ambientes de alta velocidade, como as transações eletrónicas nos mercados financeiros. Nestes cenários, a velocidade de decisão supera a capacidade de supervisão dos gestores, alertando, Laudon e Laudon (2022), de que é crucial a garantia de que os sistemas operam corretamente para evitar consequências negativas significativa.

1.3 *Business Intelligence* e as capacidades analíticas

BI é um termo amplamente utilizado por fornecedores de *software* e consultores de Tecnologia da Informação, para descrever uma infraestrutura que permite o armazenamento, integração, análise de dados empresariais. De acordo com Laudon e Laudon (2022), esta infraestrutura inclui tecnologias como bases de dados, armazéns de dados, *data marts*, *Hadoop* e plataformas analíticas. Os mesmos autores referem que

objetivo principal do BI é recolher, organizar e disponibilizar informações relevantes para os gestores, facilitando a tomada de decisões empresariais com base em dados verídicos e atualizados. Tvrđíková (2007) salienta que as ferramentas de BI, além de melhorarem a qualidade dos dados, também facilitam uma gestão mais eficaz, promovendo uma confiança acrescida por parte dos utilizadores e impulsionando a inovação organizacional. De acordo com Alasiri e Salameh (2020), os Sistemas de Apoio à Decisão como parte do BI, melhoram significativamente a eficácia da tomada de decisões ao integrar e processar grandes volumes de dados provenientes de múltiplas fontes. De acordo com os mesmos autores, a eficácia do BI nas organizações depende da sua capacidade de fornecer informações detalhadas e contextuais que, em conjunto com os Sistemas de Apoio à Decisão, resultam numa vantagem competitiva ao aumentar a capacidade de previsão e resposta às mudanças no mercado.

O impacto positivo do BI estende-se também ao desempenho organizacional. A integração bem-sucedida de BI com os processos empresariais, conforme descrito por Rouhani et al. (2016), traduz-se em benefícios organizacionais tangíveis, como o aumento do controlo organizacional e a melhoria da eficiência operacional. Os autores referem ainda que estes sistemas, além de otimizarem a análise e o planeamento estratégico, permitem uma maior agilidade na resposta às condições de mercado, conforme demonstrado pelos seus efeitos na melhoria do desempenho global das empresas.

1.3.1 Integração de BI com Tecnologias Emergentes

A evolução tecnológica acelerada tem transformado o panorama de BI, expandindo as suas capacidades e aplicabilidade em várias indústrias. Com o surgimento de tecnologias como computação na nuvem (ou *cloud computing*) e a inteligência artificial, a integração destas com BI tornou-se uma tendência estratégica para organizações que procuram otimizar as suas operações e a tomada de decisões (Laudon & Laudon, 2022).

A computação em nuvem oferece diversos benefícios que contribuem para a escalabilidade do BI. Plataformas de BI baseadas na nuvem permitem às organizações aceder e analisar grandes volumes de dados em tempo real, sem as limitações de infraestruturas físicas. Além disso, a *cloud* facilita a partilha e integração de dados entre diferentes departamentos e localizações geográficas, o que é essencial em ambientes

empresariais globais. Segundo Laudon e Laudon (2022), as soluções de BI baseadas na *cloud* permitem uma maior flexibilidade, agilidade e uma redução significativa nos custos de implementação e manutenção.

A integração da inteligência artificial, em particular, da aprendizagem computacional (ou *machine learning*), com o BI está a revolucionar a capacidade de previsão e automatização de decisões. Segundo Laudon e Laudon (2022), os modelos de aprendizagem computacional podem ser treinados para analisar padrões complexos de dados e prever resultados futuros, com uma precisão que supera métodos tradicionais. Além disso, sistemas de BI apoiados por inteligência artificial são capazes de automatizar o processo de geração de relatórios, recomendações e alertas, reduzindo o esforço humano e aumentando a velocidade da resposta organizacional. Por exemplo, sistemas de BI equipados com inteligência artificial podem identificar padrões de comportamento do cliente e sugerir automaticamente ajustes em estratégias de *marketing* ou vendas.

1.3.2 Melhores Práticas para uma Implementação Eficaz de BI

Para garantir o sucesso na implementação de sistemas de BI, algumas práticas são consideradas essenciais, nomeadamente as descritas por Singh (2023) e que se apresentam seguidamente.

- **Alinhamento Estratégico:** Antes da implementação de BI, é crucial que as empresas definam claramente os seus objetivos estratégicos e garantam que as soluções de BI estão alinhadas com esses objetivos. A falta de alinhamento entre a estratégia de BI e os objetivos de negócio pode resultar em desperdício de recursos e esforços mal direcionados.
- **Gestão da Mudança:** Para minimizar a resistência à mudança, é importante envolver todas as partes interessadas desde o início do processo, garantindo que compreendem os benefícios que o BI trará para a organização. Programas de formação e desenvolvimento são fundamentais para promover a aceitação e utilização das ferramentas de BI.
- **Foco na Qualidade dos Dados:** A implementação de uma estrutura sólida de *data governance* deve ser uma prioridade. Garantir que os dados são precisos,

consistentes e acessíveis é fundamental para que as análises produzam informações importantes.

- **Integração com Outras Tecnologias:** A combinação de BI com outras tecnologias, como computação na nuvem e inteligência artificial, pode amplificar os benefícios obtidos. A adoção de soluções de BI na nuvem, por exemplo, permite uma escalabilidade superior e uma redução de custos operacionais.
- **Avaliação Contínua de Desempenho:** Para garantir que as soluções de BI estão a gerar o retorno esperado, é fundamental implementar métricas e indicadores de desempenho que permitam avaliar regularmente o impacto dessas soluções no desempenho organizacional.

Desta forma, os sistemas de BI oferecem plataformas bem desenvolvidas não só para a gestão de dados, mas também para suportar decisões estratégicas e operacionais, garantindo que as empresas possam responder de forma mais eficiente às necessidades dinâmicas do mercado Singh (2023).

1.4 Qualidade dos dados

À medida que o volume de dados continua a crescer, especialmente com a proliferação de *Big Data*, é imperativo que as organizações implementem práticas de gestão para garantir a qualidade dos dados e, desta forma, manter a credibilidade dos seus dados em níveis elevados (Heinrich et al., 2017). *Big Data* refere-se a um volumoso conjunto de dados, frequentemente semiestruturados ou não estruturados, que torna difícil a sua análise através dos meios de processamento tradicionais (Laudon & Laudon, 2022).

A qualidade dos dados é um dos aspetos mais críticos para o sucesso de qualquer sistema de apoio à decisão. Sem dados precisos, coerentes e relevantes, os sistemas de BI tornam-se suscetíveis a erros, o que pode comprometer a capacidade das organizações em tomar decisões informadas. De facto, a má qualidade dos dados pode resultar em decisões incorretas, custos adicionais e perda de competitividade (Heinrich et al., 2017).

A qualidade dos dados é frequentemente definida em termos da sua precisão, completude, consistência, atualidade, acessibilidade e relevância para o contexto em que são utilizados. De acordo com Rubin (2014), a qualidade dos dados é multifacetada e inclui aspetos como a objetividade, credibilidade e veracidade, sendo esta última uma dimensão

crucial que se refere à fiabilidade e à verdade dos dados analisados. O mesmo autor identifica a veracidade como o "quarto V" do *Big Data*, complementando os já conhecidos volume, velocidade e variedade. Este conceito foca-se na "veracidade", ou seja, a confiança que se pode depositar nos dados disponíveis. A veracidade lida com a precisão, a exatidão, a consistência e a fiabilidade dos dados, bem como a sua origem e autenticidade. Num mundo onde os dados são gerados a partir de diversas fontes – redes sociais, sensores, plataformas digitais –, garantir a veracidade dos dados é um desafio cada vez mais relevante.

Impacto da Baixa Qualidade dos Dados: Os dados de baixa qualidade representam um risco significativo para as organizações, resultando em análises imprecisas e decisões erradas. Segundo estudos citados por Vikash e Kumar (2023), a má qualidade dos dados custa às empresas norte-americanas cerca de 3,1 mil milhões de dólares por ano. Este impacto financeiro é uma prova clara de que os dados imprecisos podem ter consequências desastrosas, desde a afetação incorreta de recursos até falhas no atendimento ao cliente e na previsão de tendências de mercado. Nos sistemas de apoio à decisão, a má qualidade dos dados pode resultar em relatórios e *dashboards* enganosos, o que compromete a capacidade dos gestores em monitorizar com precisão o desempenho organizacional. Além disso, a confiança dos utilizadores nos sistemas de BI é diretamente afetada pela qualidade dos dados apresentados. Dados inconsistentes ou desatualizados levam à desconfiança e à redução do uso de sistemas de análise de dados como suporte à decisão.

A Veracidade e a Incerteza dos Dados: A veracidade está fortemente ligada à incerteza nos dados. Como observado por Rubin (2014), os dados de grandes volumes são muitas vezes imprecisos e incompletos. As ambiguidades e os erros nos dados podem ser intencionais, como no caso de desinformação, ou involuntários, resultantes de falhas na recolha e processamento dos dados. Além disso, a incerteza pode resultar da subjetividade inerente ao conteúdo dos dados, especialmente em plataformas onde opiniões e sentimentos são expressos. A veracidade dos dados requer uma gestão adequada da incerteza, que pode ser feita através de técnicas como a fusão de dados (combinação de várias fontes para aumentar a precisão) e o uso de modelos matemáticos avançados, como por exemplo a lógica difusa.

Veracidade dos Dados no Contexto do *Big Data*: No contexto do *Big Data*, a veracidade dos dados é particularmente difícil de assegurar devido à natureza heterogénea e, muitas vezes, não estruturada dos dados recolhidos de várias fontes. Os dados provenientes de redes sociais, por exemplo, podem estar repletos de desinformação ou ruído, dificultando a sua utilização em análises críticas de negócios. A má qualidade dos dados, nestas plataformas, tem aumentado a necessidade de desenvolver novos métodos e ferramentas para avaliar a sua fiabilidade e credibilidade, antes que possam ser integrados em sistemas de apoio à decisão. Uma das soluções propostas é o uso de aprendizagem computacional e técnicas de análise semântica para avaliar a veracidade dos dados, identificando padrões de inconsistência, desinformação e subjetividade nos conteúdos textuais. Estas abordagens oferecem uma forma de quantificar a qualidade dos dados, permitindo que as empresas identifiquem e eliminem fontes de dados pouco confiáveis, aumentando assim a qualidade das suas análises (Vikash & Kumar, 2023).

Métricas para a Qualidade de Dados: Para garantir a qualidade dos dados, as organizações precisam de adotar métricas claras e eficazes. Existem várias abordagens para medir a qualidade dos dados, como detalhado por Rubin (2014), que propõem um índice de veracidade que mede a objetividade, a verdade e a credibilidade dos dados. Este índice é uma forma prática de avaliar a qualidade dos dados, especialmente em grandes volumes de dados textuais, onde a incerteza do conteúdo e da expressão pode comprometer a sua utilização para análise. Além disso, os autores Heinrich et al. (2017) referem que outras métricas comuns incluem:

- **Precisão:** O grau em que os dados refletem corretamente a realidade.
- **Consistência:** A ausência de contradições nos dados entre diferentes sistemas ou relatórios.
- **Completude:** O nível em que todos os dados necessários estão disponíveis.
- **Atualidade:** O quão recentes são os dados e se são atualizados em tempo hábil para o uso.

Desafios na Melhoria da Qualidade dos Dados: Melhorar a qualidade dos dados não é uma tarefa simples e envolve vários desafios. De acordo com Vikash e Kumar (2023), um dos maiores desafios é a diversidade das fontes de dados. Com o crescente uso de plataformas digitais, as empresas têm de lidar com dados de diferentes formatos e níveis

de qualidade. Além disso, a falta de definições padronizadas de veracidade e de metodologias consistentes para a avaliação da qualidade dos dados dificulta a implementação de soluções universais. Outro desafio reside na escalabilidade das soluções para melhorar a qualidade dos dados. Com o aumento exponencial no volume de dados, as soluções tradicionais de limpeza e validação de dados tornam-se insuficientes. Sendo assim, novas ferramentas baseadas em inteligência artificial e algoritmos avançados são necessárias para lidar com a complexidade e a escala dos dados na atualidade.

Portanto, a qualidade dos dados é essencial para garantir a eficácia dos sistemas de apoio à decisão. Sem dados precisos, completos e consistentes, as empresas arriscam tomar decisões com base em informações erradas, o que pode levar a perdas financeiras significativas e à perda de competitividade no mercado (Heinrich et al., 2017).

1.5 Boas práticas de desenvolvimento de relatórios e *storytelling*

Um design eficaz em *dashboards* e relatórios para gestão organizacional facilita o processo de tomada de decisões, ao permitir que utilizadores compreendam rapidamente dados complexos Martins et al. (2022). O *storytelling* é uma técnica que utiliza elementos narrativos para apresentar dados de forma que ajude o público a compreender as implicações dos mesmos (Dhanoa et al., 2024). Os mesmos autores reforçam que o uso de *storytelling* interativo é fundamental para retirar indicadores de forma envolvente, oferecendo uma estrutura que guia os utilizadores pela análise e aumenta o impacto dos relatórios. Ao aplicar essas práticas, os relatórios tornam-se ferramentas estratégicas essenciais no apoio à decisão.

Estrutura e Clareza na Apresentação: A clareza visual é um dos componentes essenciais no desenvolvimento de relatórios. Os autores Martins et al. (2022) destacam que a organização lógica dos dados e o uso de um layout estruturado ajudam os utilizadores a identificar rapidamente informações relevantes. Uma apresentação clara e objetiva pode ser alcançada por meio de *dashboards* organizados e fáceis de navegar. Além disso, é importante evitar o excesso de informação para não sobrecarregar o utilizador. Uma das melhores práticas é garantir que cada visualização de dados tenha um propósito claro, atendendo diretamente às questões principais que o relatório visa

responder. Focar-se nos indicadores mais relevantes torna os relatórios mais fáceis de compreender e mais úteis para a tomada de decisão.

Utilização de *Dashboards* Interativos: Os *dashboards* interativos são ferramentas eficazes para a apresentação de relatórios dinâmicos. A interatividade permite aos utilizadores personalizar a experiência de análise de dados, ajustando filtros e parâmetros para explorar diferentes camadas de informação. Os *dashboards* interativos também aumentam a flexibilidade na apresentação, proporcionando uma visão geral ou acesso a métricas mais detalhadas, adaptando os relatórios a diferentes necessidades das audiências e níveis de análise (Dhanoa et al., 2024).

Uso de *Storytelling* para Comunicar Dados: É importante guiar os utilizadores através dos dados para garantir que consigam entender as informações apresentadas. Uma boa narrativa de dados começa com uma introdução clara, onde se destacam os principais resultados e métricas de desempenho. A partir daí, os dados são apresentados de forma sequencial, levando o utilizador numa jornada lógica que explica as causas e os efeitos das tendências observadas. Esta abordagem não só facilita a interpretação dos dados, como também torna a análise mais envolvente, permitindo que os utilizadores se conectem melhor com as informações apresentadas (Dhanoa et al., 2024).

Adoção de Melhores Práticas de *Design*: Para garantir que os relatórios sejam eficazes, as melhores práticas de design devem ser aplicadas de forma estratégica, abrangendo desde a escolha de visualizações até a estruturação dos elementos no *dashboard*. Os autores Martins et al. (2022) destacam que o design de *dashboards* deve equilibrar usabilidade e estética para criar uma interface que seja tanto funcional quanto visualmente atraente. Segundo os mesmos autores, um layout bem estruturado ajuda a reduzir o esforço cognitivo dos utilizadores ao navegar pelo relatório, permitindo-lhes identificar informações relevantes de maneira rápida e intuitiva. Uma maneira eficaz de alcançar esse equilíbrio é utilizar uma estrutura hierárquica que posicione os KPI principais no topo do *dashboard*, com informações mais detalhadas em áreas inferiores, seguindo o fluxo natural de leitura de cima para baixo.

Os autores Martins et al. (2022) também ressaltam a importância de se considerar o Diagrama de Gutenberg ao projetar a interface do relatório. Esse modelo sugere que o

canto superior esquerdo de uma página é geralmente o ponto de foco inicial dos utilizadores, então é recomendável posicionar indicadores críticos nessa área. Isso facilita a orientação visual dos utilizadores e assegura que as métricas mais importantes estejam acessíveis logo de início.

Outro aspeto fundamental do *design*, segundo Numerro (2022), é a simplicidade visual. Limitar a quantidade de informações em cada página do relatório evita o excesso de dados e reduz o risco de sobrecarga visual. O autor recomenda que o número de indicadores exibidos seja mantido entre seis e dez por página, o que permite que os utilizadores assimilem os dados de forma mais clara e objetiva. Assim, o relatório torna-se mais focado nos principais objetivos, e os utilizadores conseguem se concentrar nos pontos mais importantes sem distrações.

A consistência visual é outra prática essencial. Usar um esquema de cores padronizado e manter formatos consistentes em datas, métricas e unidades ajudam os utilizadores a interpretar os dados de forma mais rápida. Numerro sugere a adoção de uma paleta de até seis cores para evitar ruído visual e garantir que o *dashboard* tenha uma aparência coesa. Essa prática é especialmente importante quando os relatórios são partilhados entre diferentes equipas ou departamentos, pois promove uma experiência de utilizador mais harmoniosa e intuitiva.

Importância da Personalização: Cada grupo de utilizadores tem necessidades diferentes quando se trata de relatórios. Assim, a personalização do conteúdo e das visualizações de dados é uma prática essencial para garantir que os relatórios são úteis para todos os tipos de utilizadores. A ferramenta de BI deve ser suficientemente flexível para permitir que os relatórios sejam facilmente adaptados, devendo oferecer opções de filtragem e segmentação de dados para que cada utilizador possa explorar as informações mais relevantes para o seu trabalho. Esta capacidade de ajustar os relatórios conforme o perfil do utilizador aumenta a eficácia da comunicação e garante que os dados são utilizados da forma mais eficiente possível (Wexler et al., 2017).

Este processo de melhoria contínua garante que os relatórios não se tornam obsoletos com o tempo e permanecem uma ferramenta essencial para o processo da tomada de decisão.

1.6 *Power BI* e modelos semânticos

O *Power BI* é um *software* de BI desenvolvido pela *Microsoft*, amplamente utilizado para transformar dados em indicadores estratégicos através de relatórios interativos e dinâmicos. No contexto empresarial, o *Power BI* destaca-se pela sua capacidade de relacionar e visualizar grandes volumes de dados provenientes de diversas fontes, permitindo que as organizações tomem decisões informadas e baseadas em dados (Microsoft Learn, 2024). Um modelo semântico é uma estrutura conceptual que define as relações entre os elementos de dados dentro de um conjunto de dados. Dentro da arquitetura do *Power BI*, os modelos semânticos desempenham um papel crucial, pois atuam como uma camada lógica que simplifica a complexidade dos dados, integrando as suas relações e cálculos de forma centralizada. Estes modelos semânticos não organizam apenas a informação, mas também permitem uma interpretação coerente dos dados, assegurando que os utilizadores têm uma visão unificada e consistente dos mesmos, independentemente das suas origens (Bansal, 2023).

No *Power BI*, existem três ferramentas de visualização dos dados que suportam o processo de desenvolvimento dos modelos semânticos, sendo estas: *Data View*, *Report View* e *Model View*. Cada uma destas ferramentas de visualização oferece funcionalidades específicas (Kaushik, 2024).

***Data View*:** A *Data View* é utilizada para importar diferentes fontes de dados, transformando e estruturando os dados para análise. A *Data View* exhibe todos os dados importados, facilitando a observação e edição dos mesmos, através de funcionalidades de ordenação, agrupamento e combinação de dados. Esta vista é fundamental para preparar os dados antes de os utilizar nas visualizações, assegurando que estão corretamente estruturados para análises precisas (Kaushik, 2024).

No *Power BI*, a *Data View* inclui a ferramenta *Power Query*, uma ferramenta para transformação e preparação de dados que simplifica o processo de extração, transformação e carregamento (ETL, do inglês *Extract, Transform, Load*). Através do Editor de *Power Query*, os utilizadores conseguem importar diversas fontes de dados e aplicar transformações de forma intuitiva, sem necessidade de programação, já que o *Power Query* gera automaticamente o código necessário para cada transformação. Com

mais de 350 tipos de transformações e conectividade a centenas de fontes de dados, o *Power Query* facilita a limpeza e a transformações dos dados, permitindo atualizações automáticas e uma experiência consistente tanto *online* como em *desktop* (Microsoft Learn, 2024).

Model View: A *Model View* oferece uma visão geral das relações entre as tabelas e entidades de dados, representando estas conexões através de diagramas UML que ilustram as relações e os atributos dos dados fornecidos. Esta ferramenta de visualização é essencial para compreender e gerir as relações entre as tabelas, garantindo que as ligações entre os dados estão corretamente definidas para análises e visualizações precisas (Kaushik, 2024). Na *Model View*, é utilizado o Modelo Entidade-Relacionamento (ER), que organiza os dados em tabelas relacionadas. O modelo ER estrutura as informações com base em entidades (ou seja, tabelas que representam diferentes conjuntos de dados) e relacionamentos (as conexões entre estas tabelas) (Microsoft Learn, 2023).

As entidades são compostas por tabelas de dimensão e de transações. As tabelas de dimensão descrevem entidades de negócio (como produtos, clientes ou tempo) e contêm informações descritivas para categorizar e agrupar dados, com uma coluna-chave que identifica cada entidade. A tabela de datas é uma dimensão comum para análises temporais. Por outro lado, as tabelas de transações registam eventos ou transações, como vendas e pedidos, e incluem colunas de medidas (ex.: quantidades ou receitas) e colunas-chave que as relacionam com tabelas de dimensão. A granularidade das tabelas de transações, ou seja, o nível de detalhe, é determinada pelas colunas de dimensão associadas. As tabelas de dimensão geralmente contêm poucos registos, enquanto as tabelas de transações podem ter milhões de linhas e expandem-se ao longo do tempo, exigindo uma estruturação eficiente para melhorar os modelos semânticos (Microsoft Learn, 2023).

Desta forma, para escolher o modelo ER a usar, é necessário considerar o volume dos dados, assim como as várias tabelas a relacionar. O *Star Schema* é uma estrutura de modelos ER popular em BI, onde uma tabela de transações central se relaciona diretamente a várias tabelas de dimensões, formando um formato de estrela. Esta abordagem é simples e eficiente, permitindo consultas rápidas, uma estrutura intuitiva e suporte a agregações (Microsoft Learn, 2023).

O *Snowflake Schema* é uma variação do modelo em estrela, em que algumas tabelas de dimensão são divididas em várias tabelas relacionadas para aumentar a organização e a normalização. Embora reduza redundâncias, este modelo tende a tornar as consultas mais lentas e a estrutura mais complexa para os utilizadores, sendo recomendado apenas quando o volume de dados o justifica (Microsoft Learn, 2023).

Esta abordagem permite integrar dados de várias fontes num único modelo unificado, essencial para a criação dos relatórios. Os Modelos ER facilitam a criação de hierarquias, filtros e medidas, sendo um dos elementos fundamentais para garantir a consistência e a integridade dos dados ao longo de diferentes visualizações e relatórios. No *Power BI*, o uso de modelos ER simplifica o processo de atualização dos dados, uma vez que todas as análises e visualizações associadas ao modelo ajustam-se automaticamente às alterações nos dados (Microsoft Learn, 2023).

Report View: A *Report View* é a principal ferramenta de criação de visualizações no *Power BI*, permitindo aos utilizadores transformarem conjuntos de dados em visuais interativos (Kaushik, 2024). As exibições de visualizações usadas para monitorizar e auxiliar na compreensão dos dados entendem-se por *dashboards*. Segundo Wexler et al. (2017), uma das principais vantagens do *Power BI* é a possibilidade de criar *dashboards* interativos que apresentam uma vasta gama de gráficos, tabelas e relatórios personalizáveis, permitindo uma análise visual eficaz de indicadores chave de desempenho (KPI, do inglês *Key Performance Indicators*).

Desta forma, os utilizadores podem criar colunas calculadas, medidas e KPI utilizando o DAX, uma linguagem específica do *Power BI* para cálculos avançados. O DAX é um dos componentes mais poderosos no *Power BI*. É uma linguagem de fórmulas desenvolvida pela *Microsoft* para criar cálculos personalizados e medidas dinâmicas. O DAX permite que os utilizadores executem operações avançadas, como agregações condicionais, cálculos temporais e manipulação de dados. Seguem-se funções das medidas em DAX comumente utilizadas (Cotton, 2022):

CALCULATE: Esta função é usada para criar cálculos condicionais, alterando filtros em medidas ou colunas.

REMOVEFILTERS: Remove filtros de uma ou mais colunas/tabelas no contexto da função *CALCULATE*, permitindo realizar cálculos ignorando esses filtros específicos.

SAMEPERIODLASTYEAR: Devolve um conjunto de datas no mesmo período do ano anterior. É útil para comparações anuais, como vendas do mesmo mês no ano passado.

TOTALYTD (Year-to-Date): Calcula o valor acumulado desde o início do ano até a data atual.

MAXX: Avalia uma expressão para cada linha de uma tabela e devolve o valor máximo resultante. Esta função é usada para encontrar o maior valor baseado em cálculos.

SWITCH: Avalia uma expressão em múltiplas condições e devolve valores diferentes dependendo da condição que for atendida. Funciona como uma “fórmula de escolha” para múltiplos casos.

SUMX: Executa uma soma sobre uma expressão calculada para cada linha de uma tabela. Esta função é útil para somas condicionais e cálculos mais complexos.

USERELATIONSHIP: Ativa uma relação inativa entre duas tabelas para um cálculo específico. Esta função é usada em contextos com múltiplas relações entre tabelas.

TOPN: Retorna as primeiras N linhas de uma tabela, ordenadas de acordo com uma expressão. Desta forma, obtém-se o “top” de elementos selecionados.

SUMMARIZE: Cria uma tabela onde se especifica as colunas e as medidas que são necessárias de incluir.

A interatividade dos *dashboards* do *Power BI* permite que os utilizadores personalizem as visualizações de acordo com as suas necessidades específicas, filtrando dados, ajustando intervalos de tempo e explorando detalhes mais profundos com um simples clique. Isso facilita a exploração e interpretação de grandes volumes de dados de forma prática e eficiente (Wexler et al., 2017).

Os modelos semânticos no *Power BI* promovem a consistência e a integridade dos dados, uma vez que centralizam cálculos, medidas e relações que podem ser partilhadas entre múltiplos relatórios. Com uma camada semântica bem estruturada, as organizações conseguem reduzir a complexidade associada à análise de dados, garantindo que as equipas têm acesso a dados fiáveis e bem organizados (Bansal, 2023).

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na *Zeone Consulting*

Através do *Power BI Service*, plataforma *online* que suporta a partilha e a colaboração de relatórios criados no *Power BI Desktop*, as organizações conseguem publicar os seus relatórios, tornando-os acessíveis para diferentes utilizadores de acordo com as permissões atribuídas. Uma funcionalidade fundamental para a segurança dos dados é a Segurança ao Nível de Linha (RLS, do inglês *Row Level Security*), que permite restringir o acesso a informações sensíveis com base no perfil do utilizador. Com a RLS, cada utilizador visualiza apenas os dados que lhe dizem respeito, garantindo que informações confidenciais são protegidas e que a experiência de análise é personalizada para cada perfil (Microsoft Learn, 2024).

Por exemplo, numa organização onde diferentes equipas acedem ao mesmo relatório, a RLS pode ser configurada para mostrar apenas os dados relevantes para cada equipa ou nível hierárquico. Isto assegura uma maior governação dos dados, uma vez que as permissões são geridas centralmente e aplicadas de forma consistente em todos os relatórios associados ao modelo semântico. Além de proporcionar segurança, a RLS também otimiza a eficiência ao reduzir a necessidade de criar diferentes relatórios para diferentes utilizadores, concentrando toda a análise num único modelo seguro e adaptado (Microsoft Learn, 2024).

2 *Zeone Consulting*

Este capítulo apresenta a *Zeone Consulting* e o seu papel de suporte no Grupo FHC, descrevendo a organização, serviços principais e visão de inovação da empresa. Também aborda o contexto do estágio no Departamento de Planeamento Estratégico, destacando o desenvolvimento de modelos semânticos em *Power BI* e as práticas de confidencialidade e integridade dos dados tratados.

2.1 Apresentação

Inserida no Grupo FHC, a *Zeone Consulting* é uma empresa especializada em tecnologias da informação, desempenhando um papel crucial como suporte corporativo para as demais empresas do grupo, como apresentado na Figura 2.1. De acordo com informação interna da empresa, a *Zeone* posiciona-se como um parceiro estratégico, proporcionando soluções tecnológicas inovadoras que são fundamentais para o sucesso e a eficiência das operações do Grupo FHC.

O Grupo FHC é um dos mais destacados grupos empresariais no setor farmacêutico em Portugal, com origens e capitais exclusivamente privados portugueses. Presente em quatro continentes e em mais de 65 países, o Grupo FHC desenvolve atividades que cobrem toda a cadeia de valor do medicamento, incluindo áreas industriais, distribuição no mercado ambulatorio e hospitalar, exportação, internacionalização, investigação e desenvolvimento (I&D) em saúde, bem como tecnologias de informação (Group FHC, 2024).

A missão do Grupo FHC é clara: contribuir ativamente para melhorar o acesso a cuidados de saúde de qualidade em todo o mundo. Com mais de 25 anos de experiência, o grupo tem-se consolidado como um ator fundamental no setor da saúde, oferecendo produtos e serviços que respondem às necessidades não atendidas de pacientes e consumidores a nível global (Group FHC, 2024).



Figura 2.1- Posicionamento da *Zeone Consulting* no Grupo FHC

Visão e Organização por Processos: A visão da *Zeone* é ser uma referência no mercado de consultoria em tecnologias da informação, através da oferta de serviços que não só otimizam processos empresariais, mas também ajudam as empresas a adaptarem-se às novas exigências tecnológicas. A organização da *Zeone* é centrada em processos bem definidos que envolvem as seguintes etapas: reunir informações, planear, desenhar soluções, desenvolver, testar e, finalmente, entregar os projetos. Este enfoque por processos permite à *Zeone* garantir qualidade, inovação e eficácia em todos os serviços prestados (*Zeone Consulting*, 2024).

Serviços Oferecidos pela *Zeone Consulting*: A *Zeone* oferece uma vasta gama de serviços, entre os quais se destacam (*Zeone Consulting*, 2024):

- Desenvolvimento de *Software*: Personalizado para atender às necessidades específicas de cada cliente, assegurando que as soluções implementadas são eficientes e alinhadas com os objetivos do negócio.
- Consultoria Financeira e Auditoria: A *Zeone* apoia os seus clientes na gestão dos aspetos financeiros, assegurando o cumprimento das normas e otimização dos resultados através de auditorias rigorosas e serviços de assessoria financeira.
- Formação e Gestão de Pessoas: Reconhecendo a importância do capital humano, a *Zeone* desenvolve programas de formação e projetos de gestão de talento, contribuindo para a melhoria contínua das competências das equipas das empresas clientes.

- **Assessorias Contabilística, Fiscal e Jurídica:** A *Zeone* oferece suporte em áreas críticas como a contabilidade, fiscalidade e questões jurídicas, garantindo que as empresas operam dentro dos requisitos legais e maximizam as suas vantagens fiscais.

Futuros Projetos e Expansão: De acordo com informação interna da empresa, a *Zeone Consulting* está empenhada em continuar a sua expansão, tanto no desenvolvimento de novas soluções tecnológicas como no fortalecimento das suas parcerias estratégicas. A empresa planeia investir em projetos que alavanquem a digitalização e a automatização dos processos empresariais, mantendo-se na vanguarda das tecnologias emergentes. Este compromisso com a inovação garante que a *Zeone* permanecerá como um pilar essencial no suporte corporativo do Grupo FHC, contribuindo para o crescimento sustentado e o sucesso global do grupo.

2.2 Estágio

Inserido no Departamento de Planeamento Estratégico da *Zeone Consulting*, o estágio teve como objetivo principal contribuir para a melhoria contínua dos processos de tomada de decisão da empresa. Este departamento é composto por uma equipa de três pessoas, cada uma com responsabilidades específicas e complementares, que colaboram para garantir a qualidade e a relevância das informações estratégicas.

O departamento é liderado pelo responsável pela coordenação das atividades e pela supervisão dos projetos de análise de dados. Neste departamento, são desenvolvidos os modelos semânticos em *Power BI*, criando uma estrutura lógica e organizada que facilita a criação de relatórios interativos. Para assegurar que os dados provenientes de SQL são importados em *Power BI* já estruturados, existe um elemento do departamento responsável pelo processo de ETL em SQL. Desta forma, é necessário apenas realizar ajustes pontuais em casos excecionais.

Através deste trabalho conjunto, o departamento consegue garantir que os dados utilizados nos modelos semânticos são precisos e consistentes, proporcionando uma base sólida para a tomada de decisão estratégica.

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na *Zeone Consulting*

É importante salientar que, na qualidade de estagiária na *Zeone Consulting*, todos os dados utilizados neste trabalho foram provenientes exclusivamente da empresa, garantindo assim a confidencialidade e a integridade das informações tratadas. O tratamento dos dados foi realizado com rigorosos cuidados para garantir a confidencialidade e a integridade das informações da empresa. Como estagiária, há um compromisso adicional de respeito pelas políticas de confidencialidade, para assegurar que nenhuma informação sensível seja divulgada ou manipulada indevidamente.

3 Metodologia do trabalho

Neste capítulo será apresentada a metodologia CRISP-DM, assim como sua adaptação ao contexto das atividades realizadas no estágio.

3.1 Metodologia CRISP-DM

A metodologia escolhida para o trabalho desenvolvido foi adaptada da metodologia CRISP-DM que consiste num conjunto de fases e processos padrões para desenvolver projetos de DM. Desenvolvido em 1996 por um consórcio de empresas, como DaimlerChrysler e SPSS, e apresentado no documento Chapman et al. (2000), que se segue de perto nesta subsecção, o CRISP-DM foi concebido para ser uma metodologia não proprietária, flexível e aplicável a diferentes setores e ferramentas de DM. Como apresentado na Figura 3.1, o modelo está estruturado em seis fases principais que abrangem desde a compreensão do negócio até à implantação dos resultados.

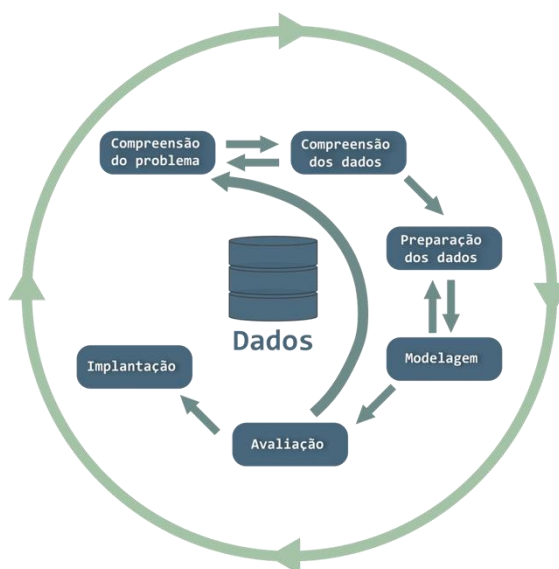


Figura 3.1 - Fases do ciclo de vida da metodologia CRISP-DM.

Fonte: Adaptado de Chapman et al. (2000)

Compreensão do negócio: A fase inicial de um projeto de DM centra-se na compreensão dos objetivos e requisitos do negócio. O principal objetivo é alinhar o processo de DM com os objetivos estratégicos da organização (Chapman et al., 2000).

Compreensão dos dados: Nesta fase, são abrangidas a exploração e a descrição dos dados. Este entendimento inicial permite também identificar subconjuntos de dados interessantes para uma análise mais aprofundada (Chapman et al., 2000).

Preparação dos dados: Nesta fase, os dados são preparados para serem utilizados na modelação. Isto inclui a seleção das variáveis relevantes, a limpeza de dados inconsistentes ou incorretos e a integração de diferentes fontes de dados (Chapman et al., 2000).

Modelação: A fase de modelação consiste na seleção e aplicação de técnicas de DM e desenvolvimento de modelos de representação do conhecimento descoberto, assim como a avaliação do desempenho dos algoritmos escolhidos. Dependendo do problema de negócio, podem ser aplicadas técnicas como regressão, *clustering* ou classificação (Mishra et al., 2017).

Avaliação: Após a construção dos modelos, estes são avaliados com base em critérios previamente definidos para garantir que satisfazem os objetivos de negócio estabelecidos. A avaliação dos modelos assegura que os mesmos respondem às perguntas de negócio e que os resultados obtidos são aplicáveis e úteis.

Implantação: A fase final do CRISP-DM é a implantação, onde os resultados da DM são integrados nos processos de negócio da organização. Nesta fase, os resultados da modelação são aplicados no ambiente real de negócio, com planos para manutenção e monitorização contínua dos modelos.

3.2 Metodologia aplicada ao contexto do estágio

Uma vez que não foram usadas técnicas de DM, as fases da metodologia CRISP-DM foram adaptadas para o contexto do estágio e os trabalhos realizados. Assim, as fases entendem-se por:

Análise e entendimento do negócio: Esta fase inicial centrou-se na compreensão dos objetivos e requisitos do negócio. Esta etapa envolveu a definição clara das questões de

negócio a serem respondidas e a análise do contexto empresarial, para garantir que as ações futuras estivessem alinhadas com os objetivos do projeto. A análise cuidada dos recursos, restrições e com diferentes condições foi fundamental para planear adequadamente os diferentes modelos semânticos e evitar esforços desnecessários numa direção errada.

Seleção e entendimento dos dados: Após a compreensão dos objetivos de negócio, a fase seguinte consistiu em explorar os dados disponíveis. Nesta fase, os dados foram analisados quanto à sua qualidade e adequação ao problema a ser resolvido. A avaliação da completude, consistência e precisão dos dados foi crucial, uma vez que os problemas de qualidade podem comprometer a eficácia das fases subsequentes dos modelos semânticos.

Preparação dos dados: No contexto do estágio, a preparação dos dados passou por utilizar a ferramenta de transformação de dados no *Power BI*, o *Power Query*. Esta ferramenta desempenhou um papel importante na transformação de dados, permitindo realizar operações como a remoção de duplicados, o tratamento de valores em falta e a deteção de anomalias. O *Power Query* permitiu ainda tarefas como a divisão e fusão de colunas de texto, de modo a estruturar e limpar a informação de forma eficiente. O objetivo foi garantir que os dados estivessem prontos para serem usados na fase seguinte (Microsoft Learn, 2024).

Modelação: Nesta fase, foram desenvolvidos os modelos semânticos correspondentes a cada conjunto de dados. Assim, começou-se por desenvolver modelos ER, criaram-se medidas em DAX e foram contruídas representações visuais das interligações entre diferentes elementos de dados. A ferramenta *Model View* do *Power BI* permitiu integrar múltiplas fontes de dados e organizar os dados de forma eficiente, facilitando a análise das relações entre dimensões e factos através de esquemas, como o esquema estrela, amplamente utilizado para organizar os modelos ER, e o esquema floco de neve para estruturas de dados mais complexas. Esta flexibilidade permitiu uma melhor interpretação e usabilidade dos dados, de modo a melhorar a compreensão dos mesmos e ajudar a gerar melhores indicadores (Microsoft Learn, 2023).

Avaliação dos modelos: No contexto do estágio, a fase da avaliação dos modelos foi uma etapa essencial antes da sua implantação. Verificou-se se as medidas DAX devolviam os resultados esperados, fazendo testes de validação para garantir a qualidade dos dados apresentados. Além disso, outro ponto crucial foi assegurar que a segurança dos dados, como a segurança ao nível de linhas ou a segurança ao nível de objetos está devidamente implementada para impedir o acesso não autorizado aos relatórios resultantes dos modelos semânticos desenvolvidos. Caso os resultados da avaliação não fossem satisfatórios, seria necessário regressar às fases anteriores para ajustar o modelo semântico ou alterar dados (Microsoft Learn, 2024).

Implantação: Por fim, procedeu-se à implantação dos modelos semânticos desenvolvidos, onde os resultados foram integrados nos processos de negócio da organização. Nesta fase, cada modelo semântico desenvolvido nas fases anteriores foi aplicado ao contexto real e publicado na plataforma *online Power BI Service*. Esta etapa foi importante para transformar e ajustar os processos de acordo com os objetivos e o *feedback* dos utilizadores. A implantação garantiu também que o sistema incluísse mecanismos de atualizações e manutenções contínua, de forma a assegurar que os modelos continuassem a gerar valores fidedignos ao longo do tempo, permitindo uma melhoria contínua na gestão das operações (Microsoft Learn, 2024).

4 Desenvolvimento de Modelos Semânticos

Como mencionado na Secção 2.2, no decorrer do estágio realizado na *Zeone Consulting*, todos os dados provieram de fontes transversais às várias empresas do Grupo FHC. Durante o este período, dois pilares fundamentais orientaram os trabalhos desenvolvidos: a garantia da qualidade dos dados apresentados e a aplicação de boas práticas no uso do *Power BI*.

4.1 Enquadramento das atividades desenvolvidas nas fases do CRISP-DM

A qualidade dos dados foi assegurada através de processos rigorosos em cada etapa do projeto, desde a preparação dos dados até à apresentação final, garantindo que os dados estivessem corretos, consistentes e prontos para serem utilizados nas análises. Além disso, a implementação de boas práticas no *Power BI* foi essencial para garantir que os relatórios fossem intuitivos, funcionais e alinhados com as necessidades de cada área de negócio, contribuindo para uma tomada de decisão mais eficaz.

Ao longo deste percurso, diferentes conjuntos de dados foram utilizados, cada um exigindo níveis distintos de intervenção nas várias fases do desenvolvimento dos modelos semânticos correspondentes.

A Tabela 4.1 define o enquadramento dos modelos semânticos trabalhados nas fases adaptadas do CRISP-DM. Cada fase, desde a análise do negócio, seleção e preparação de dados, até à modelação, avaliação e implantação, foram cuidadosamente adaptadas às características e necessidades dos dados, assegurando a correta preparação e tratamento para garantir a qualidade das análises e resultados apresentados.

Tabela 4.1 - Fases adaptadas do CRISP-DM usadas em cada modelo semântico.

| | 1ª Fase | 2ª Fase | 3ª Fase | 4ª Fase | 5ª Fase | 6ª Fase |
|------------------------|---------|---------|---------|---------|---------|---------|
| <i>Stocks (Orders)</i> | X | X | | | X | |
| <i>Worktime</i> | X | X | | | X | |

| | | | | | | |
|---------------------------|---|---|---|---|---|---|
| <i>People Cost</i> | X | X | X | X | X | X |
| <i>Fluorinated Gases</i> | X | X | | X | X | X |
| <i>Chemical Products</i> | X | X | X | X | X | X |
| <i>Strategic Planning</i> | X | X | X | X | X | X |
| <i>Utilities</i> | X | X | X | X | X | X |

Na fase inicial de integração na empresa, os modelos “*Stocks (Orders)*” e “*Worktime*” foram utilizados para compreender o funcionamento dos dados na organização. Estes permitiram realizar a análise do negócio e dos dados, o que facilitou a avaliação dos modelos com o objetivo de definir, desde o início, os principais critérios para a validação e qualidade dos dados.

O modelo “*Fluorinated Gases*” foi essencial para entender de que forma os modelos semânticos de dados e as boas práticas no desenvolvimento de relatórios influenciavam a perceção visual dos utilizadores no momento da consulta dos relatórios. A análise deste modelo possibilitou observar como a apresentação dos dados impactava a interpretação das informações por parte dos utilizadores.

Com os conhecimentos adquiridos na análise dos modelos semânticos anteriores, foi possível desenvolver os restantes modelos, como “*People Cost*”, “*Chemical Products*”, “*Strategic Planning*” e “*Utilities*”, de acordo com os dois pilares fundamentais: a garantia da qualidade dos dados e a adoção de boas práticas no desenvolvimento de relatórios interativos, garantindo que os modelos e os relatórios respondiam às necessidades do negócio.

4.2 Análise e entendimento do negócio

Na primeira fase do estágio, conforme o modelo CRISP-DM, foi realizada uma análise focada no entendimento do negócio e no contexto em que a *Zeone Consulting* se insere no Grupo FHC. Sob a orientação do supervisor, foi recomendada a consulta de alguns

modelos semânticos já existentes, permitindo compreender a organização dos dados e de que forma o departamento fornecia os relatórios aos utilizadores.

Este processo foi crucial para obter uma visão clara da estrutura da empresa e das suas operações, especialmente no que diz respeito à gestão de dados e disponibilização de relatórios. A análise dos modelos existentes ajudou a perceber os fluxos de trabalho e a identificar as principais necessidades de informação dos utilizadores. A partir deste entendimento, foi possível alinhar o projeto de análise de dados com os objetivos e expectativas do departamento, garantindo que os resultados obtidos pudessem contribuir para a otimização dos processos de negócio e para a tomada de decisões estratégicas no grupo.

Nos modelos semânticos criados posteriormente, o foco esteve em compreender o tema específico de cada um deles e como estes se inseriam na estrutura e operações da empresa.

4.3 Seleção e entendimento dos dados

A segunda fase do modelo CRISP-DM envolveu a compreensão dos dados que seriam utilizados no processo de análise. Nesta fase, foi essencial entender detalhadamente os conjuntos de dados com que se iria trabalhar, identificando as informações relevantes e eliminando os dados que não contribuía diretamente para os objetivos da análise.

A empresa utilizava várias fontes de dados, nomeadamente *SharePoint*, *Dataverse* e *SQL*, sendo que o entendimento dos dados variava consoante a sua origem. Embora os dados provenientes de *SQL* fossem importados no *Power Query* já preparados, ainda assim exigiam uma análise cuidadosa para compreender como as diferentes tabelas se iriam relacionar. Neste caso, foi necessário identificar as tabelas de dimensão e tabelas de transação para garantir que os dados selecionados permitissem futuramente a correta relação entre todas as tabelas nos modelos semânticos correspondentes.

Independentemente da fonte de dados, era sempre necessário perceber quais os dados que facilitariam a criação de um modelo ER coerente e eficiente. Nos conjuntos de dados provenientes de *SharePoint* e *Dataverse*, foi necessária uma verificação de todas as colunas, visto que parte das colunas tinham de ser excluídas ou ajustadas posteriormente para garantir a sua adequação ao modelo semântico.

Esta fase permitiu assegurar que os dados estavam prontos para serem utilizados de forma eficaz nas fases subsequentes, garantindo que as relações entre as tabelas fossem otimizadas para suportar as necessidades de análise do negócio.

4.4 Preparação dos dados

A fase de preparação de dados foi uma das mais importantes no processo de análise, garantindo que os dados estivessem prontos para serem utilizados na modelação e subsequente análise. Nesta fase, foram adotadas diferentes abordagens conforme a fonte dos dados, com destaque para os conjuntos de dados provenientes de SQL, *SharePoint* e *Dataverse*.

Os modelos semânticos que passaram por esta fase de preparação foram: "*Human Resources*", "*Chemical Products*", "*Strategic Planning*" e "*Utilities*". Nestes modelos, foram realizados ajustes importantes para garantir a qualidade e consistência dos dados, permitindo que fossem adequadamente utilizados no modelo ER.

Em todos os modelos semânticos, a primeira etapa da preparação de dados foi a importação da tabela de datas que serviu como tabela de dimensão das datas. Também foi necessária a criação de parâmetros que definissem o período que a tabela de datas deveria apresentar valores, como apresentado na Figura 4.1:

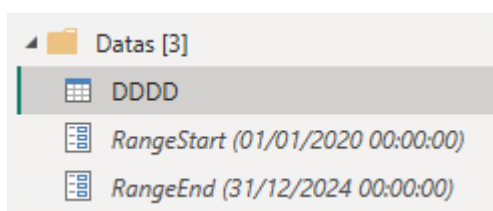


Figura 4.1- Tabela de datas e parâmetros correspondentes

Nos conjuntos de dados provenientes de SQL, a maior parte da preparação já tinha sido feita. No entanto, foi necessário especificar o tipo de dados de algumas colunas. Exemplos comuns incluíam:

- Colunas de datas, que eram frequentemente importadas como “data e tempo”, como observado na Figura 4.2, e precisavam ser convertidas para o tipo "data", como apresentado na Figura 4.3.

| | DataAdmissao | |
|--|---------------------|---|
| | 01/03/2012 00:00:00 | M |
| | 01/03/2012 00:00:00 | M |
| | 01/03/2011 00:00:00 | F |
| | 01/03/2011 00:00:00 | F |
| | 01/03/2011 00:00:00 | M |

Figura 4.2- Coluna "DataAdmissao" no formato original

| | DataAdmissao | |
|---|--------------|--|
| ? | 01/03/2012 | |
| ? | 01/03/2012 | |
| ? | 01/03/2011 | |
| ? | 01/03/2011 | |
| ? | 01/03/2011 | |

Figura 4.3- Coluna "DataAdmissao" após alteração do tipo de data

- Colunas numéricas inteiras, por vezes importadas como texto, que necessitavam de conversão para o formato numérico apropriado.
- Foi necessário a criação de novas colunas calculadas, como apresentado na Figura 4.4, e posteriormente, proceder à normalização dos dados:

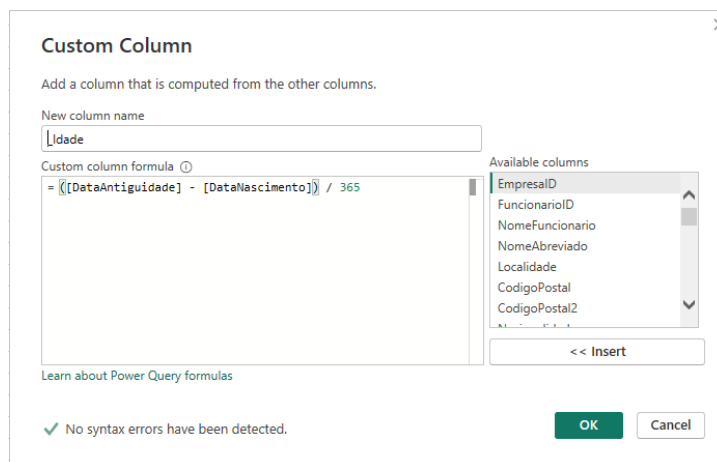


Figura 4.4- Criação de coluna calculada em *Power Query*

| ABC 123 | _Idade |
|------------|---------------------|
| | 56.13:24:49.3150684 |
| | 52.13:32:42.7397260 |
| | 44.16:57:51.7808219 |
| | 38.07:49:28.7671232 |
| | 48.08:17:05.7534246 |
| | 41.05:03:46.8493150 |
| | 43.11:02:47.6712328 |
| | 34.23:44:13.1506849 |

Figura 4.5- Coluna "_Idade" como resultado da criação de coluna calculada

| 1 ² 3 | _Idade |
|---------------------|--------|
| | 56 |
| | 52 |
| | 44 |
| | 38 |
| | 48 |
| | 41 |
| | 43 |
| | 34 |
| | 39 |

Figura 4.6- Coluna "_Idade" após alteração de tipo de dados

A Figura 4.5 e a Figura 4.6 ilustram o processo de conversão de tipos de dados no *Power Query* da coluna “_Idade”, onde foi necessário obter a idade de cada colaborador. Desta forma, a coluna calculada obtida na Figura 4.5, no formato de dados geral, foi ajustada para o formato correto, como demonstrado pela Figura 4.6.

Nos conjuntos de dados provenientes de *SharePoint* e *Dataverse*, a preparação foi mais extensa e incluiu:

- Remoção de colunas desnecessárias: Filtragem de colunas que não eram relevantes para o modelo de dados, uma vez que a importação das tabelas vem sempre com colunas desnecessárias, como ilustrado pela Figura 4.7.

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

49 COLUMNS, 307 ROWS Column profiling based on top 1000 rows

Figura 4.7- Tabela "ZZUtilitieRubrica" após importação em *Power Query*

Desta forma, uma tabela com 49 colunas, passa a ser uma tabela com apenas 6 colunas, como pode ser observado na Figura 4.8.

6 COLUMNS, 307 ROWS Column profiling based on top 1000 rows

Figura 4.8- Tabela "ZZUtilitieRubrica" após remoção de colunas desnecessárias

- Eliminação de dados duplicados: Identificação e remoção de registos duplicados para assegurar a integridade das análises.
- Expansão de colunas: Em algumas tabelas, foi necessário expandir colunas que continham valores aninhados, apresentado na Figura 4.9 e na Figura 4.10.

| zeo_artigos(zeo_artigos) |
|--------------------------|
| Value |
| Value |
| Value |
| Value |
| Value |
| Value |
| Value |
| Value |

Figura 4.9- Coluna "zeo_artigos(zeo_artigos)" após importação em Power Query

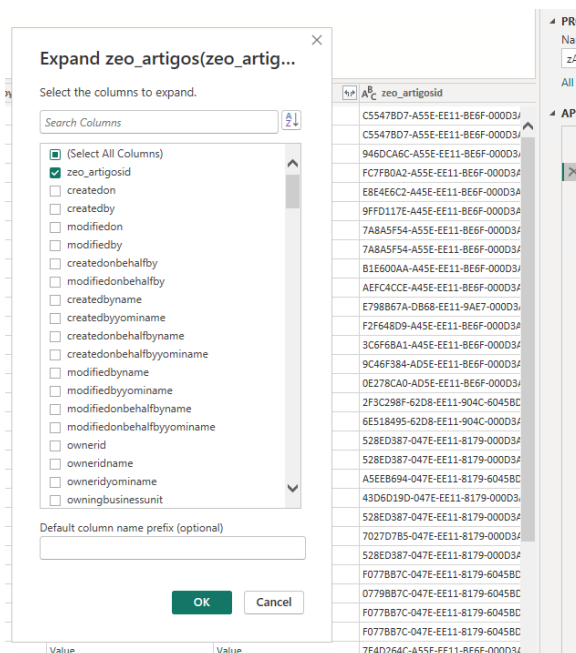


Figura 4.10- Coluna "zeo_artigos(zeo_artigos)" expandida

- Normalização dos dados: Garantiu-se a consistência dos dados para facilitar a criação de um modelo ER. Como apresentado pela Figura 4.11 e pela Figura 4.12, esta normalização implicava aplicar diversas transformações às tabelas de modo a obter uma tabela normalizada.

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

The screenshot shows a Power Query table with the following data:

| EmpresID | PlantName | PlantAddress | Latitude | Longitude | Active | PlantName - Copy |
|----------|---------------------|------------------------------------------------------------------|-----------|-----------|--------|------------------|
| 1 | GRUPO FHC-BAS-LSM | Portugal ContinentalMortiguaMortigua Portuga3450 | 40.391751 | -8.183789 | TRUE | BAS |
| 2 | GRUPO FHC-FHC-UL1a2 | Portugal ContinentalMortiguaMortigua Portuga3450 | 40.392377 | -8.184186 | TRUE | FHC |
| 3 | GRUPO FHC-FHC-UL3 | Portugal ContinentalMortiguaMortigua Portuga3450 | 40.392544 | -8.184378 | TRUE | FHC |
| 4 | GRUPO FHC-EMP-MOV | Portugal ContinentalMontemor-o-VelhoMontemor-o-Velho Portuga3... | 40.193537 | -8.641169 | TRUE | EMP |
| 5 | GRUPO FHC-BAS-UM | Portugal ContinentalMortiguaMortigua Portuga3450 | 40.391943 | -8.176214 | TRUE | BAS |

Figura 4.11- Coluna "PlantName - Copy" com valores extraídos entre delimitadores

The screenshot shows a Power Query table with the following data:

| PlantName | PlantAddress | Latitude | Longitude | Active | PlantName - Company | Unit |
|---------------------|------------------------------------------------------------------|-----------|-----------|--------|---------------------|-----------|
| GRUPO FHC-BAS-LSM | Portugal ContinentalMortiguaMortigua Portuga3450 | 40.391751 | -8.183789 | TRUE | BAS | LSM |
| GRUPO FHC-FHC-UL1a2 | Portugal ContinentalMortiguaMortigua Portuga3450 | 40.392377 | -8.184186 | TRUE | FHC | UL1 UL2 |
| GRUPO FHC-FHC-UL3 | Portugal ContinentalMortiguaMortigua Portuga3450 | 40.392544 | -8.184378 | TRUE | FHC | UL3 |
| GRUPO FHC-EMP-MOV | Portugal ContinentalMontemor-o-VelhoMontemor-o-Velho Portuga3... | 40.193537 | -8.641169 | TRUE | EMP | MOV |
| GRUPO FHC-BAS-UM | Portugal ContinentalMortiguaMortigua Portuga3450 | 40.391943 | -8.176214 | TRUE | BAS | UM |

Figura 4.12- Coluna "Unit" com alteração de valores

- Criação de colunas calculadas e condicionais: A criação de colunas calculadas e condicionais em *Power Query* foi importante para enriquecer os dados brutos e prepará-los para a análise. Muitas vezes, os dados originais não continham toda a informação necessária, e estas colunas permitiram realizar cálculos ou transformações que ajudaram a organizar os dados, como ilustrado pela Figura 4.13. Desta forma, assegurou-se que o modelo semântico seria mais eficiente.

Add Conditional Column ×

Add a conditional column that is computed from the other columns or values.

New column name

| Column Name | Operator | Value | Output |
|-------------|---------------------|--------|--------------------------------------------|
| If | zeo_tpoderiscona... | equals | ABC 123 Advertências de Perigo - Patrim... |
| Then | ABC 123 | Sim | ... |

Else

Figura 4.13- Criação de coluna calculada condicional "*Património Genético*"

Em contraste, os modelos semânticos "*Stocks (Orders)*", "*Worktime*" e "*Fluorinated Gases*" já se encontravam em fases mais avançadas, e, por isso, não necessitaram de preparação de dados.

Independentemente da fonte dos dados, o objetivo final desta fase foi garantir que os dados estivessem devidamente tratados para suportar a criação de um modelo ER eficaz. Isso envolveu identificar as tabelas de dimensão e transação e assegurar que as relações entre elas estavam bem definidas.

4.5 Modelação

A quarta fase do modelo CRISP-DM focou-se na construção dos modelos semânticos, de forma a organizar e estruturar a informação proveniente dos diferentes conjuntos de dados. Os modelos semânticos que passaram por este processo de foram: "*Human Resources*", "*Chemical Products*", "*Strategic Planning*", "*Utilities*" e "*Fluorinated Gases*".

4.5.1 Modelos ER

Para os modelos que passaram por esta fase, foi necessário criar relações claras e bem definidas entre tabelas de dimensão e transação, permitindo que a análise cruzada entre diferentes categorias de dados fosse realizada de forma fluida e precisa. Já os modelos

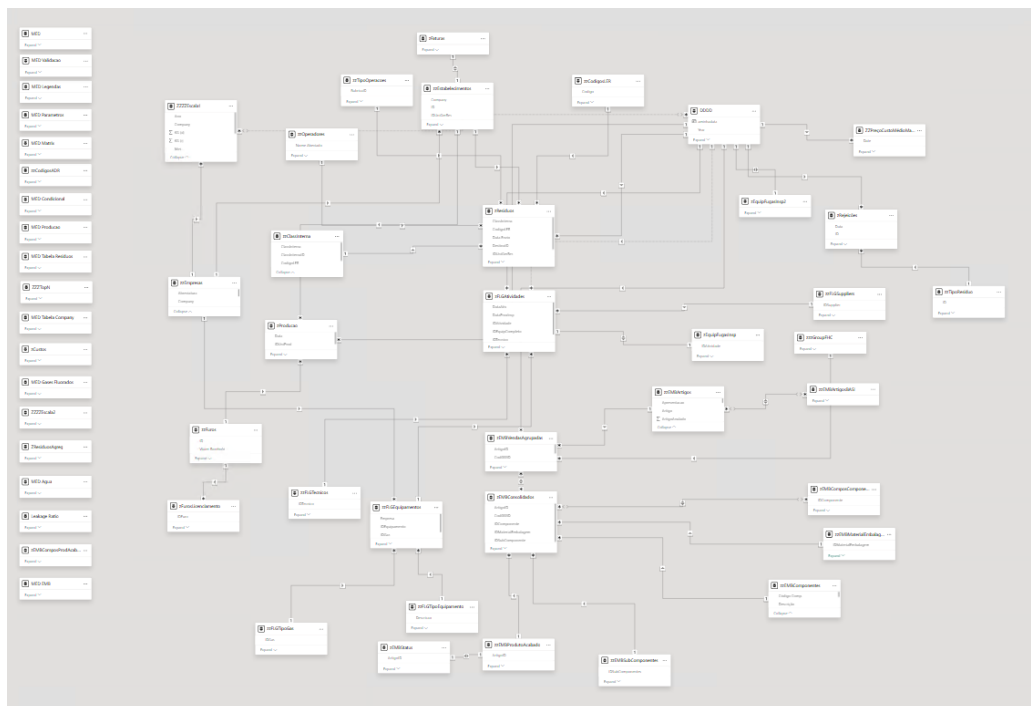


Figura 4.15- Modelo ER: *Snowflake Schema*

A Figura 4.15 corresponde a um modelo *snowflake schema*, este apresenta uma estrutura mais detalhada, na qual algumas dimensões são normalizadas e desdobradas em várias tabelas relacionadas. Esta abordagem permitiu uma maior organização dos dados e evitou-se redundâncias, apesar de ser mais complexa em comparação com o modelo *star schema*. Ambos os esquemas permitiram a criação de relatórios interativos, com análises dinâmicas baseadas em diferentes parâmetros de consulta, adaptando-se às necessidades específicas de cada análise.

4.5.2 Medidas em DAX

Além da construção do modelo ER, foram criadas várias medidas em DAX para ampliar a capacidade de análise dos dados. Entre as medidas DAX criadas, destacaram-se cálculos como somas de valores, percentagens e cálculos acumulados, que permitiram um desenvolvimento de modelos semânticos aprofundados. As funções DAX mais utilizadas no desenvolvimento dessas medidas foram essenciais para realizar cálculos dinâmicos e interativos, permitindo trabalhar com os dados de forma eficaz. Seguidamente, são listadas as principais funções utilizadas.

- *CALCULATE* com múltiplos filtros:

```

1 Total Salary Cost =
2 CALCULATE(
3     SUM(zProcessamentos[__CustosFuncionarios])
4 )

```

Figura 4.16- Medida "Total Salary Cost"

A medida apresentada na Figura 4.16 utiliza a função *CALCULATE* para somar todos os valores da coluna “__CustosFuncionarios” da tabela “zProcessamentos”. O objetivo é obter o custo total com os funcionários, isto é, a soma de todos os valores de custos associados a salários nessa tabela.

```

1 Total Consumos N =
2 var yyyyy = SELECTEDVALUE(DDDD[Year])
3 var indicador = SELECTEDVALUE(ZZZIndicadoresEner[Descricao])
4
5 var vab =
6 CALCULATE(
7     SUM( ZZVAB[VAB]
8         , USERRELATIONSHIP(ZZEmpUnidades[Facility], ZZVAB[Unidade])
9         , DDDD[Year] = yyyyy
10 )
11 var intensidade_ener = CALCULATE(
12     [TEPs consumo total]*1000/vab
13     , DDDD[Year] = yyyyy
14 )
15 var cons_ener = CALCULATE(
16     [TEPs consumo total]*1000/'MED Produção'[Production Final]
17     , DDDD[Year] = yyyyy
18 )
19 var cons_ener_unit = CALCULATE(
20     ([TEPs consumo total]*1000)/[Production Units 2]
21     , DDDD[Year] = yyyyy
22 )
23 var intensidade_carb = CALCULATE(
24     [Ton CO2 (SGCIE) total] / [TEPs consumo total]
25     , DDDD[Year] = yyyyy)
26 var ind_valor =
27 SWITCH(
28     TRUE()
29     , indicador = "Intensidade Carbónica" , intensidade_carb
30     , indicador = "Intensidade Energética" , intensidade_ener
31     , indicador = "Consumo Especifico de Energia" , cons_ener
32     , indicador = "Consumo Especifico de Energia (Unidades)" , cons_ener_unit
33     , BLANK()
34 )
35
36 RETURN
37 SWITCH(
38     TRUE()
39     , yyyyy <= 2020 , ""
40     , ind_valor)
41
42

```

Figura 4.17- Medida "Total Consumos N"

```

1 Total Consumos N-1 =
2 var yyyy = SELECTEDVALUE(DDDD[Year]) - 1
3 var indicador = SELECTEDVALUE(ZZZIndicadoresEner[Descricao])
4
5 var vab =
6 CALCULATE(
7     SUM( ZZVAB[VAB]
8         , USERRELATIONSHIP(ZZEmpUnidades[Facility], ZZVAB[Unidade])
9         , DDDD[Year] = yyyy
10    )
11 var intensidade_ener = CALCULATE(
12     [TEPs consumo total]*1000/vab
13     , DDDD[Year] = yyyy
14 )
15 var cons_ener = CALCULATE(
16     [TEPs consumo total]*1000/'MED Produção'[Production Final]
17     , DDDD[Year] = yyyy
18 )
19 var cons_ener_unit = CALCULATE(
20     ([TEPs consumo total]*1000)/[Production Units 2]
21     , DDDD[Year] = yyyy
22 )
23 var intensidade_carb = CALCULATE(
24     [Ton CO2 (SGCIE) total] / [TEPs consumo total]
25     , DDDD[Year] = yyyy
26 )
27 var ind_valor =
28 SWITCH(
29     TRUE()
30     , indicador = "Intensidade Carbónica" , intensidade_carb
31     , indicador = "Intensidade Energética" , intensidade_ener
32     , indicador = "Consumo Específico de Energia" , cons_ener
33     , indicador = "Consumo Específico de Energia (Unidades)" , cons_ener_unit
34     , BLANK()
35 )
36
37 RETURN
38 SWITCH(
39     TRUE()
40     , yyyy <= 2020 , ""
41     , ind_valor)
42

```

Figura 4.18- Medida "Total Consumos N-1"

As medidas da Figura 4.17 e da Figura 4.18 calculam diferentes indicadores de consumo energético ou intensidade carbónica com base no ano selecionado e no indicador selecionado. O cálculo varia conforme o indicador selecionado, que pode ser: "Intensidade Carbónica", "Intensidade Energética", "Consumo Específico de Energia", ou "Consumo Específico de Energia por Unidades". As medidas utilizam os dados de consumo em *TEPs* e *CO2*, *VAB* e produção para calcular o indicador correspondente. Se o ano for anterior ou igual a 2020, não devolve valores. Ambas as medidas seguem a mesma lógica, mas focam-se em anos diferentes: a medida da Figura calcula os consumos totais de cada indicador para o ano selecionado e a medida da Figura calcula os consumos totais para o ano anterior. Estas medidas foram criadas com o objetivo de serem apresentadas simultaneamente de modo a ser possível comparar a variação de valores de um ano para o outro.

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

```

1  Varia Indicadores =
2  var yyyyy = SELECTEDVALUE(DDDD[Year])
3  var indicador = SELECTEDVALUE(ZZZIndicadoresEner[Descricao])
4  var vab =
5  CALCULATE(
6  SUM( ZZVAB[VAB]
7  --, FILTER(ODDD , DDDD[Year])
8  , USERRELATIONSHIP(ZZEmpUnidades[Facility], ZZVAB[Unidade])
9  , DDDD[Year] = yyyyy
10 )
11 var intensidade_ener = CALCULATE(
12 [TEPs consumo total]*1000/vab
13 , DDDD[Year] = yyyyy
14 )
15 var cons_ener = CALCULATE(
16 [TEPs consumo total]*1000/'MED Produção'[Production Final]
17 , DDDD[Year] = yyyyy
18 )
19 var cons_ener_unit = CALCULATE(
20 ([TEPs consumo total]*1000)/[Production Units 2]
21 , DDDD[Year] = yyyyy
22 )
23 var intensidade_carb = CALCULATE(
24 [Ton CO2 (SGCIE) total] / [TEPs consumo total]
25 , DDDD[Year] = yyyyy)
26 var ind_valor =
27 SWITCH(
28 TRUE()
29 , indicador = "Intensidade Carbónica" , (intensidade_carb - [Intensidade Carbónica 20])/[Intensidade Carbónica 20]
30 , indicador = "Intensidade Energética" , (intensidade_ener - [Intensidade Energética 20])/[Intensidade Energética 20]
31 , indicador = "Consumo Especifico de Energia" , (cons_ener - 'MED Consumos'[Consumo Especifico de Energia 20]) / [Consumo Especifico de Energia 20]
32 , indicador = "Consumo Especifico de Energia (Unidades)" , (cons_ener_unit - 'MED Consumos'[Consumo Especifico de Energia Unidades 20]) / 'MED Consumos'[Consumo Especifico de
33 Energia Unidades 20]
34 , BLANK()
35 )
36 RETURN
37 SWITCH(
38 TRUE()
39 , yyyyy <= 2020 , ""
40 , ind_valor)
41

```

Figura 4.19- Medida "*Varia Indicadores*"

Na Figura 4.19, a medida calcula a variação percentual de vários indicadores de desempenho energético em relação ao ano de referência, que é 2020. Dependendo do indicador selecionado, a medida devolve a variação para os diferentes indicadores.

A variação é calculada como a diferença entre o valor do indicador no ano atual (ano selecionado) e o valor de 2020, dividida pelo valor de 2020. Assim como as medidas apresentadas anteriormente, se o ano selecionado for anterior ou igual a 2020, não são devolvidos os valores correspondentes.

```

1 Headcount =
2 VAR yyyy= YEAR(TODAY())
3 var max_month_number = MAX(DDDD[Month])
4 var ano = [SelectedYear]
5 var yyyy2 = SELECTEDVALUE(DDDD[Year], BLANK())
6 var max_month = CALCULATE(SELECTEDVALUE(DDDD[Month]), DDDD[Month] = max_month_number)
7
8 VAR headcount_month =
9 CALCULATE(
10     SUM(zHeadcount[Count])
11     , zHeadcount[CodMov] <> "ZZZ_000"
12 )
13 VAR headcount_no_month =
14 IF(
15     yyyy = yyyy2
16     , CALCULATE (
17         SUM (zHeadcount[Count])
18         , zHeadcount[CodMov] <> "ZZZ_000"
19         , zHeadcount[last month] = 1
20         )
21     , CALCULATE (
22         SUM(zHeadcount[Count])
23         , zHeadcount[CodMov] <> "ZZZ_000"
24         , YEAR(zHeadcount[Data]) = ano
25         , MONTH(zHeadcount[Data]) * 1 = 12
26         )
27 )
28 var headcount_multiple_month =
29 CALCULATE(
30     SUM(zHeadcount[Count])
31     , zHeadcount[CodMov] <> "ZZZ_000"
32     , MONTH(zHeadcount[Data]) = max_month )
33
34 RETURN
35
36 SWITCH(
37     TRUE()
38     , ISFILTERED(DDDD[Month]) = FALSE() , headcount_no_month
39     , HASONEFILTER(DDDD[Month]) = TRUE(), headcount_month
40     , headcount_multiple_month
41 )

```

Figura 4.20- Medida "Headcount"

A medida apresentada na Figura 4.20 calcula o número de funcionários com base em variáveis que têm em conta o ano atual, o mês máximo de referência e o ano selecionado no relatório. A lógica da medida adapta-se às diferentes condições de filtro de tempo, utilizando diferentes abordagens para o cálculo do número de funcionários dependendo do mês ou ano selecionado.

- REMOVEFILTERS

```

1 Intensidade Energética 20 =
2 CALCULATE(
3     [Intensidade Energética (Kg)]
4     , REMOVEFILTERS(DDDD[Year])
5     , DDDD[Year] = 2020
6 )

```

Figura 4.21- Medida "Intensidade Energética 20"

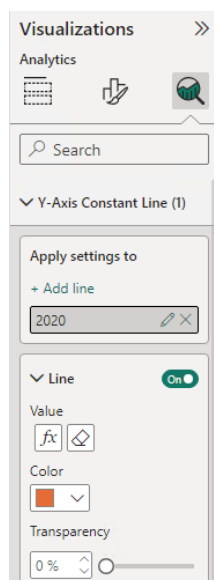


Figura 4.22- *Power BI*: Eixo Y Constante

A medida da Figura 4.21, começa por apresentar a função *CALCULATE* que é usada para calcular a medida “*Intensidade Energética (Kg)*”. Esta medida representa a intensidade energética em quilogramas, ajustando o contexto do cálculo para focar-se exclusivamente no ano de 2020.

A função *REMOVEFILTERS* remove quaisquer filtros aplicados à coluna do ano na tabela de datas, garantindo que o cálculo da intensidade energética para o ano de 2020 não será afetado por outros filtros temporais no relatório.

A medida devolve o valor da intensidade energética para o ano de 2020, que, como demonstrado pela Figura 4.22, é utilizado como uma linha constante no eixo y do gráfico, permitindo uma comparação visual entre o ano selecionado no relatório e o valor de referência de 2020.

- *SAMEPERIODLASTYEAR*

```

1 Total Salary Cost Year N-1 =
2 CALCULATE(
3     SUM(zProcessamentos[__CustosFuncionarios])
4     , SAMEPERIODLASTYEAR(DDDD[aminhadata])
5 )
6

```

Figura 4.23- Medida "*Total Salary Cost N-1*"

Na Figura 4.23, a medida calcula o custo total com salários dos funcionários no ano anterior, utilizando a função *CALCULATE* em conjunto com *SAMEPERIODLASTYEAR*. A função *CALCULATE* realiza a soma dos valores da coluna “__CustosFuncionarios” na tabela “zProcessamentos”. O *SAMEPERIODLASTYEAR* ajusta automaticamente o contexto temporal para o mesmo período do ano anterior, com base na coluna de datas.

Isto permite obter o total dos custos da empresa correspondentes ao salário do ano anterior, facilitando comparações com o ano atual ou outros períodos no relatório, sendo útil para análises de tendências e desempenho ao longo do tempo.

- *TOTALYTD (Year-to-Date)*

```

1 YTD Overtime (hrs) =
2 CALCULATE(
3     TOTALYTD([Overtime (hrs)], DDDD[aminhadata])
4 )

```

Figura 4.24- Medida "YTD Overtime (hrs)"

A medida apresentada na Figura 4.24 calcula o total de horas extras acumuladas desde o início do ano até à data atual, para cada colaborador, com o objetivo de identificar quando é que esse colaborador atinge um acumulado de 32 horas úteis. Esta contagem é fundamental para calcular o direito a descanso compensatório. A função *CALCULATE* é utilizada para alterar o contexto do cálculo, enquanto a *TOTALYTD* acumula as horas extras registadas pela medida da soma das horas extras ao longo do ano, com base na coluna de datas.

- *MAXX*

```

1 eixo intensidade energetica =
2 var IntensidadeEner = CALCULATE(MAXX(VALUES(DDDD[Month]), [Intensidade Energética (Kg)], ALLSELECTED(DDDD)))
3
4 RETURN
5 SWITCH(
6     TRUE()
7     , [Intensidade Energética 20] > IntensidadeEner , [Intensidade Energética 20] * 1.4
8     , IntensidadeEner * 1.4
9 )

```

Figura 4.25- Medida "eixo intensidade energética"

Na Figura 4.25, a medida começa por calcular a variável “*IntensidadeEner*”, que utiliza a função *CALCULATE* para determinar o valor máximo da intensidade energética com base em diferentes meses. A função *MAXX* é aplicada à tabela resultante dos valores únicos dos meses. A função *ALLSELECTED* remove quaisquer filtros aplicados à tabela das datas, garantindo que o cálculo considera todo o contexto selecionado pelo utilizador no relatório. De seguida, a medida utiliza a função *SWITCH* para avaliar as diferentes condições e devolver o valor apropriado. Se o valor da Intensidade Energética no ano 2020 for superior ao valor da Intensidade Energética para o ano selecionado, o resultado será o valor da Intensidade Energética no ano 2020 multiplicado por 1,4, para obter uma margem no eixo Y. Caso contrário, o resultado será o valor da Intensidade Energética para o ano selecionado multiplicado por 1,4.

Esta medida calcula um valor ajustado da intensidade energética, baseado na comparação entre dois valores de intensidade energética. Desta forma, o eixo considera sempre o valor superior para identificar o valor máximo que o eixo Y deve apresentar. Esta lógica ajusta o eixo da intensidade energética de forma flexível, dependendo dos valores calculados, garantindo uma visualização adequada e proporcional dos dados no gráfico.

```
1 eixo valorização =
2 CALCULATE(MAXX(VALUE(DDDD[Month]), [Valorizacao]), ALLSELECTED(DDDD)) * 1.2
```

Figura 4.26- Medida "*eixo valorização*"

Uma vez que a medida apresentada na Figura 4.26 não precisa de ter em conta valores de anos anteriores, esta medida é uma versão simples da medida apresentada anteriormente.

- *SWITCH*

```
1 Consumo Ano Title =
2 VAR source = SELECTEDVALUE(ZZZBibliotecaTepsCO2[Source])
3
4 RETURN
5 SWITCH(
6     TRUE()
7     , source = "TEPs"      , "TEPs Consumption"
8     , source = "SGCIE"     , "Ton CO2 - SGCIE"
9     , source = "INERPA & CELE" , "Ton CO2 - INERPA & CELE"
10 )
```

Figura 4.27- Medida "*Consumo Ano Title*"

A medida desenvolvida na Figura 4.27 ajusta dinamicamente o título do gráfico apresentado nas próximas Figuras com base na opção selecionada pelo utilizador no filtro presente no gráfico. A medida utiliza a variável “source” que armazena o valor selecionado no relatório utilizando a função *SELECTEDVALUE*. Esta função devolve o valor único da fonte atualmente filtrada no relatório. De seguida, a função *SWITCH* é utilizada para avaliar o valor da variável “source” e devolver um título correspondente:

- Se, no filtro, a opção selecionada for “TEPs”, a medida devolve o título “TEPs Consumption”, como observado na Figura 4.28.

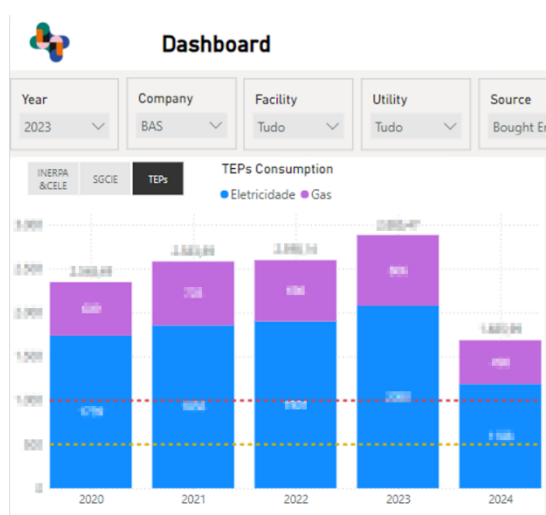


Figura 4.28- Resultado do título após selecionar “TEPs”

- Se a fonte selecionada for “SGCIE”, o título devolvido será “Ton CO2 - SGCIE”, como constatado na Figura 4.29.

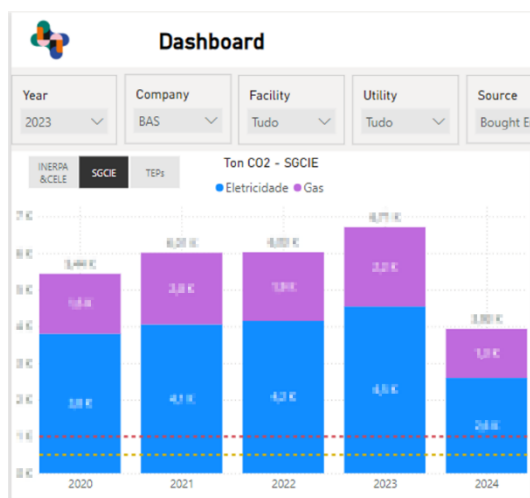


Figura 4.29- Resultado do título após selecionar "SGCIÉ"

- Se a fonte for "INERPA & CELE", o título será "Ton CO2 - INERPA & CELE", como indicado na Figura 4.30.

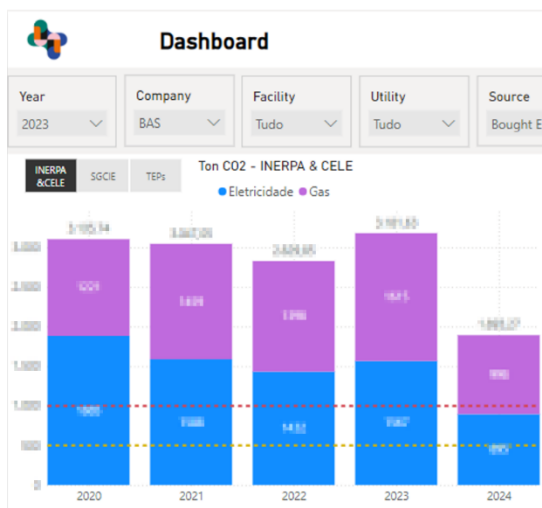


Figura 4.30- Resultado do título após selecionar "INERPA & CELE"

De seguida, as medidas apresentadas definem cores específicas para ser usada em diversas visualizações ou formatações personalizadas no *Power BI*. Uma vez que estas medidas são utilizadas em diversas medidas, em vez de criar 3 variáveis de cores para cada medida criada, criaram-se 3 medidas que são referenciadas sempre que necessário. A Figura 4.31

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

apresenta a medida de cor amarela. Na Figura 4.32, é apresentada a medida de cor verde. A medida presente na Figura 4.33, apresenta a medida de cor vermelha.

```
1 CF Amarelo = "#D9B300"
```

Figura 4.31- Medida "CF Amarelo"

```
1 CF Verde = "#3AB166"
```

Figura 4.32- Medida "CF Verde"

```
1 CF Vermelho = "#EC515C"
```

Figura 4.33- Medida "CF Vermelho"

```
1 Color Varia Indicadores =
2
3 var unid = SELECTEDVALUE(ZZEmpUnidades[Unidade])
4 var indicador = SELECTEDVALUE(ZZIIndicadoresEner[Descricao])
5
6 var cor =
7 SWITCH ( true()
8     , [Varia Indicadores] > 0.00     , [CF Vermelho]
9     , [Varia Indicadores] <= 0.00    , [CF Verde]
10    )
11
12 var ijm =
13 SWITCH ( true()
14     , [Varia Indicadores] > 0.00     , [CF Vermelho]
15     , [Varia Indicadores] <= 0.00 && [Varia Indicadores] > -0.06 , [CF Amarelo]
16     , [Varia Indicadores] <= - 0.06 , [CF Verde]
17    )
18
19 var lsm =
20 SWITCH ( true()
21     , [Varia Indicadores] > 0.00     , [CF Vermelho]
22     , [Varia Indicadores] <= 0.00 && [Varia Indicadores] > -0.04 , [CF Amarelo]
23     , [Varia Indicadores] <= - 0.04 , [CF Verde]
24    )
25
26 var outros =
27 SWITCH ( true()
28     , [Varia Indicadores] > 0.00     , [CF Vermelho]
29     , [Varia Indicadores] = 0.00     , [CF Amarelo]
30     , [Varia Indicadores] < 0.00     , [CF Verde]
31    )
32
33 RETURN
34
35 SWITCH(
36     TRUE()
37     , indicador = "Intensidade Carbónica" , cor
38     , unid = "IJM"                        , ijm
39     , unid = "LSM"                        , lsm
40     , outros
41 )
42
```

Figura 4.34- Medida "Color Varia Indicadores"

Como verificado na Figura 4.34, a medida atribui cores dinâmicas às variações de indicadores com base em condições específicas para diferentes unidades de negócio e tipos de indicadores. Utiliza várias variáveis que ajustam as cores conforme os valores da variação dos indicadores:

- Variação positiva é marcada a vermelho.
- Variação negativa é marcada a verde.
- Dependendo da unidade de negócio, valores entre intervalos específicos são marcados a amarelo para destacar uma variação moderada.

A medida utiliza a função *SWITCH* para selecionar a cor correta, garantindo uma formatação condicional que facilita a leitura da performance dos indicadores.

```

1 color waste varia n-2/n-1 =
2
3 var descri = SELECTEDVALUE(ZZZBiblioteca_RubricaResiduo[Descricao])
4
5 var all_varia =
6 SWITCH ( true()
7     , [variacao n-2/n-1] > 0     , [CF Vermelho]
8     , [variacao n-2/n-1] = 0     , [CF Amarelo]
9     , [variacao n-2/n-1] < 0     , [CF Verde]
10    )
11
12 var valorizacao_varia =
13 SWITCH ( true()
14     , [variacao n-2/n-1] > 0     , [CF Verde]
15     , [variacao n-2/n-1] = 0     , [CF Amarelo]
16     , [variacao n-2/n-1] < 0     , [CF Vermelho]
17    )
18
19 RETURN
20 SWITCH(
21     TRUE()
22     , descri = "Valorização"     , valorizacao_varia
23     , descri = "Penalidades"     , [CF Vermelho]
24     , all_varia
25 )

```

Figura 4.35- Medida "Color waste varia n-2/n-1"

Como mais um exemplo do vasto uso das medidas de cor, a medida exibida na Figura 4.35 atribui cores às variações dos resíduos entre os dois anos anteriores ao ano selecionado no relatório, com base na descrição do resíduo e no valor da variação. Existem dois comportamentos principais:

- Para a descrição "Valorização", variações positivas são verdes, variações negativas são vermelhas, e valores iguais a zero são amarelos.
- Para "Penalidades", a cor é sempre vermelha.

- Para outras descrições, variações positivas são vermelhas, variações negativas são verdes, e variações iguais a zero são amarelas.

A função *SWITCH* é utilizada para escolher a cor correta com base na variação e na descrição, permitindo uma formatação condicional visualmente clara.

- *SUMXX*

```

1 Valorizacao =
2 var valor =
3 CALCULATE(
4     SUMX(
5         zUtConsumoDetalhe
6         , IF(
7             RELATED(ZZTipoDoc[Descrição]) = "Fatura"
8             , ABS(zUtConsumoDetalhe[Total Linha])
9             , -(zUtConsumoDetalhe[Total Linha])
10        )
11    )
12    , FILTER(ZZUtilitierubrica, ZZUtilitierubrica[Utilities] = "Resíduos")
13    , FILTER(ZZUtilitierubrica, ZZUtilitierubrica[Rubrica] = "Valorização")
14 )
15 RETURN
16 valor

```

Figura 4.36- Medida "*Valorizacao*"

Na medida da Figura 4.36, a função *SUMX* percorre cada linha da tabela "*zUtConsumoDetalhe*" e calcula um valor com base em diversas condições. A função *IF* é usada para determinar se o valor da linha deve ser positivo ou negativo, dependendo do tipo de documento.

Dentro do *SUMX*, a função *IF* verifica o tipo de documento relacionado através da função *RELATED*. Se este for "Fatura", o valor da linha é convertido para um valor absoluto com a função *ABS*, garantindo que o valor seja positivo. Se não for uma "Fatura", o valor é transformado em negativo, multiplicando-o por -1.

A função *FILTER* é usada duas vezes para garantir que apenas os registos que pertencem à categoria "Resíduos" e à rubrica "Valorização" são incluídos no cálculo:

Por fim, a medida devolve o valor calculado, que é a soma ajustada dos valores das linhas (positivos ou negativos conforme o tipo de documento), exclusivamente para a rubrica de "Valorização" no contexto de "Resíduos".

- *USERELATIONSHIP*

```

1 m3 normalizado =
2 CALCULATE(
3     SUMX(
4         zUtConsumo
5         , IF(
6             RELATED(ZZTipoDoc[Descrição]) = "Fatura"
7             , zUtConsumo[m3 (n)]
8             , -(zUtConsumo[m3 (n)])
9         )
10    )
11    , USERELATIONSHIP(ZZEmpresas[Empresa], zUtConsumo[fnc_empresaid])
12 )
13
14

```

Figura 4.37- Medida "m3 normalizado"

Na Figura 4.37, a medida calcula o volume normalizado em metros cúbicos com base no tipo de documento associado e na empresa, utilizando uma relação específica entre a empresa e a tabela de consumo.

A função *SUMX* percorre cada linha da tabela “*zUtConsumo*”, calculando o valor do volume em metros cúbicos normalizado para cada linha. O resultado é acumulado ao longo das linhas da tabela.

Dentro do *SUMX*, a função *IF* verifica o tipo de documento relacionado através da função *RELATED*. Se for uma "Fatura", o valor do volume normalizado em metros cúbicos é mantido positivo. Se o documento for de outro tipo, o valor é transformado em negativo, multiplicando-o por -1.

A função *USERELATIONSHIP* ativa uma relação entre as colunas das empresas entre as tabelas “*ZZEmpresas*” e “*zUtConsumo*”. Isto permite que o cálculo seja feito especificamente para a empresa selecionada no contexto atual do relatório. Esta função é útil quando se trabalha com modelos de dados que têm relações inativas, e é necessário ativar uma relação específica durante o cálculo.

- *TOPN*

```

1 Top N ECHA =
2   CALCULATE( [Quantity (Kg) ECHA],
3
4   TOPN('ZZZTopN'[Valor de TopN], ALL(ZZArtigos[Artigo]), [Quantity (Kg) ECHA]),
5
6   VALUES(ZZArtigos[Artigo])
7 )

```

Figura 4.38- Medida "Top N ECHA"

A medida apresentada na Figura 4.38 calcula a quantidade total de um produto (em Kg) de acordo com a classificação do top N de artigos, com base na quantidade total de cada artigo. A função *TOPN* é usada para selecionar os N artigos com as maiores quantidades, onde N é definido na tabela "ZZZTopN" através da coluna "Valor de TopN". Esta tabela contém um valor que o utilizador pode selecionar para determinar o número de artigos a serem exibidos no top N. A função *TOPN* ordena os artigos pela quantidade calculada na medida indicada e seleciona os primeiros N artigos de acordo com esse valor.

Desta forma, medida "Top N ECHA" devolve a quantidade total de artigos (em Kg) que pertencem ao top N, baseado no valor de N especificado pelo utilizado. Como apresentado na Figura 4.39, se for selecionado no filtro "Top" o valor 5, vão ser devolvidos os 5 artigos com maior quantidade (em Kg). Da mesma forma, como observado na Figura 4.40, se for selecionado o valor 8, vai ser devolvido os 8 artigos com maior quantidade (em Kg). Isto permite visualizar os artigos mais relevantes em termos de quantidade, ajudando na análise e gestão dos artigos considerados como produtos químicos.

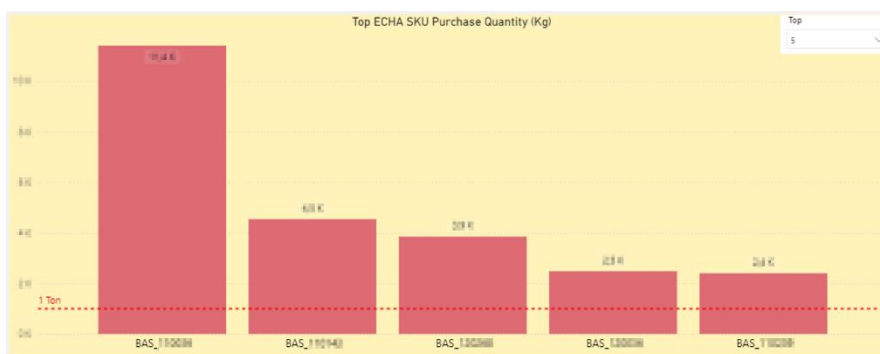


Figura 4.39- Resultado do gráfico quando selecionado o Top 5

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting



Figura 4.40- Resultado do gráfico quando selecionado Top 8

- *UNION* e *SUMMARIZE*

```

1 zHeadcount =
2 UNION(
3   CALCULATETABLE (
4     SELECTCOLUMNS (
5       SUMMARIZE (
6         zProcessamentos
7         , zFuncionarios[EmpresaID]
8         , zProcessamentos[FuncionarioID]
9         , DDDO[aminhadata]
10        , zProcessamentos[DataMOV]
11        , zProcessamentos[CodMov]
12        , zProcessamentos[__Antiguidade]
13        , zFuncionarios[DataAdmissao]
14        , zFuncionarios[DataDemissao]
15        , zFuncionarios[DataDemissao]
16        )
17        , "EmpresaID" , zFuncionarios[EmpresaID]
18        , "FuncionarioID" , zProcessamentos[FuncionarioID]
19        , "Data" , SWITCH ( TRUE(), DDDO[aminhadata] = zFuncionarios[DataAdmissao], zFuncionarios[DataAdmissao], DDDO[aminhadata] = zProcessamentos[DataMOV], zProcessamentos[DataMOV], DDDO[aminhadata] = zFuncionarios[DataDemissao], zFuncionarios[DataDemissao] )
20        , "CodMov" , zProcessamentos[CodMov]
21        , "Count" , SWITCH ( TRUE(),
22          , and (YEAR ( zFuncionarios[DataAdmissao] ) = YEAR ( zProcessamentos[DataMOV] ) , MONTH ( zFuncionarios[DataAdmissao] ) = MONTH ( zProcessamentos[DataMOV] ) ) , BLANK()
23          , and (YEAR ( zFuncionarios[DataDemissao] ) = YEAR ( zProcessamentos[DataMOV] ) , MONTH ( zFuncionarios[DataDemissao] ) <= MONTH ( zProcessamentos[DataMOV] ) ) , BLANK()
24          , 1 )
25        , "Seniority" , zProcessamentos[__Antiguidade]
26        , RIGHT ( zProcessamentos[CodMov], 3 ) = "R01"
27        )
28    )
29  )
30
31  CALCULATETABLE(
32    SELECTCOLUMNS (
33      SUMMARIZE (
34        zFuncionarios
35        , zFuncionarios[EmpresaID]
36        , zFuncionarios[FuncionarioID]
37        , zFuncionarios[DataAdmissao]
38        , zFuncionarios[DataDemissao]
39      )
40    , "EmpresaID" , zFuncionarios[EmpresaID]

```

Figura 4.41- Tabela calculada "zHeadcount" - Parte 1

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

```

29
30
31
32 CALCULATETABLE(
33     SELECTCOLUMNS (
34         SUMMARIZE (
35             ZZFuncionarios
36             , ZZFuncionarios[EmpresaID]
37             , ZZFuncionarios[FuncionarioID]
38             , ZZFuncionarios[DataAdmissao]
39             , ZZFuncionarios[DataDemissao]
40         )
41         , "EmpresaID" , ZZFuncionarios[EmpresaID]
42         , "FuncionarioID" , ZZFuncionarios[FuncionarioID]
43         , "Data" , ZZFuncionarios[DataAdmissao]
44         , "CodMov" , "ZZ_000"
45         , "Count" , SWITCH(TRUE(), and(YEAR (ZZFuncionarios[DataAdmissao]) = YEAR (ZZFuncionarios[DataDemissao]) , MONTH (ZZFuncionarios
46             [DataDemissao])) , BLANK() , 1)
47         , "Seniority" , BLANK()
48     )
49 )
50
51
52 CALCULATETABLE(
53     SELECTCOLUMNS (
54         SUMMARIZE (
55             ZZFuncionarios
56             , ZZFuncionarios[EmpresaID]
57             , ZZFuncionarios[FuncionarioID]
58             , ZZFuncionarios[DataAdmissao]
59             , ZZFuncionarios[DataDemissao]
60             , ZZFuncionarios[Antiguidade]
61         )
62         , "EmpresaID" , ZZFuncionarios[EmpresaID]
63         , "FuncionarioID" , ZZFuncionarios[FuncionarioID]
64         , "Data" , ZZFuncionarios[DataAdmissao]
65         , "CodMov" , "ZZ_000"
66         , "Count" , SWITCH(TRUE(), ZZFuncionarios[DataDemissao] = ZZFuncionarios[DataAdmissao] , BLANK() , 1)
67         , "Seniority" , ZZFuncionarios[Antiguidade]
68     )
69 )

```

Figura 4.42- Tabela calculada "zHeadcount" - Parte 2

A tabela “zHeadcount” apresentada na Figura 4.41 e na Figura 4.42 foi criada para consolidar e calcular o número de funcionários da empresa em diferentes momentos e situações, utilizando a função UNION para combinar três conjuntos de dados principais:

O primeiro conjunto de dados agrupa informações dos processamentos e funcionários, como o ID da empresa, ID do funcionário, datas relevantes (data de admissão, movimentação ou demissão), e o código de movimentação. No campo “Count”, observado na Figura 4.41, é utilizada uma lógica condicional para contar apenas os funcionários ativos no momento do processamento, ignorando as situações em que a data de admissão coincide com a de movimentação ou demissão.

Como demonstrado na Figura 4.42, o segundo conjunto de dados inclui apenas os funcionários admitidos contando somente os funcionários que não tiveram demissão no mesmo mês em que foram admitidos.

O terceiro conjunto de dados registra os funcionários demitidos, contando apenas os funcionários demitidos, a menos que a data de admissão coincida com a de demissão.

A função UNION combina estes três conjuntos de dados, permitindo criar uma visão completa do histórico de funcionários, com detalhes de movimentações, admissões e demissões.

```

1 Base salary (nominal) =
2 var vencimentos =
3 SUMMARIZE(
4     zProcessamentos
5     , zProcessamentos[FuncionarioID]
6     , zProcessamentos[DataMOV]
7     , "VencimentoBruto" , CALCULATE(AVERAGE(zProcessamentos[__Vencimento]))
8 )
9
10 var calculo_vencimento = CALCULATE(SUMX(vencimentos, [VencimentoBruto]))
11
12 RETURN
13
14 calculo_vencimento

```

Figura 4.43- Medida "*Base salary (nominal)*"

Na Figura 4.43, a medida calcula o vencimento base nominal dos funcionários. Inicialmente, cria uma variável chamada *vencimentos* que utiliza a função *SUMMARIZE* para agrupar os dados da tabela “*zProcessamentos*” com base no ID do funcionário e na data de movimentação do processamento. Dentro de cada grupo, é calculada a média dos vencimentos brutos de cada funcionário através da função *AVERAGE*, com os resultados armazenados numa nova coluna chamada "*VencimentoBruto*".

Em seguida, a variável “*calculo_vencimento*” utiliza a função *SUMX* para somar todos os valores da coluna "*VencimentoBruto*" na tabela agregada, resultando no cálculo final do vencimento base nominal agregado para todos os funcionários. Por fim, a medida devolve o valor total dos vencimentos base nominais, oferecendo uma visão clara da média dos vencimentos ao longo do tempo.

```

1 Overtime (hrs) =
2 var tabela_overtime =
3 SELECTCOLUMNS (
4     CALCULATETABLE (
5         SUMMARIZE (
6             zProcessamentos
7             , zProcessamentos[FuncionarioID]
8             , zProcessamentos[CodMov]
9             , zProcessamentos[DataMOV]
10            , zProcessamentos[Tempo]
11        )
12        , zProcessamentos[TipoTabela] = "2"
13        , zProcessamentos[DataMOV] >= 2023/01/01
14    )
15    , "FuncionarioID"          , zProcessamentos[FuncionarioID]
16    , "Data Movimento"       , zProcessamentos[DataMOV]
17    , "Cod Movimento"       , zProcessamentos[CodMov]
18    , "Tempo"                , CALCULATE(SUM(zProcessamentos[Tempo]))
19 )
20
21 var overtime =
22 CALCULATE(
23     SUMX(tabela_overtime, [Tempo])
24 )
25
26 RETURN
27 overtime

```

Figura 4.44- Medida "Overtime (hrs)"

Como observado na Figura 4.44, a medida calcula o total de horas extras realizadas por cada colaborador, filtrando os dados a partir de uma determinada data e para um tipo específico de registo de processamento.

Na medida começa-se por criar uma tabela temporária chamada “*tabela_overtime*”, utilizando a função *SELECTCOLUMNS* para definir as colunas a serem extraídas. Esta tabela é construída com base na função *CALCULATETABLE*, que resume os dados da tabela “*zProcessamentos*” através da função *SUMMARIZE*, agrupando-os por diversas colunas.

O filtro “*zProcessamentos[TipoTabela]*” garante que apenas são considerados os processamentos correspondentes a horas extras. Além disso, é aplicada uma restrição temporal para incluir apenas os dados a partir de 1 de janeiro de 2023, de maneira a limitar o volume de dados criados pela tabela temporária. Em seguida, a coluna “*Tempo*” calcula a soma das horas extras para cada grupo.

Após a construção da tabela temporária, a variável “*overtime*” utiliza a função *SUMX* para somar todas as horas extras da tabela criada temporariamente. Desta forma, obtém-se o total de horas extras realizadas pelos funcionários.

```

1 SKU from several regions =
2
3 var Tabela_SKU = SUMMARIZE ( zArtigosComprasDetalhe
4 | | | | | , zArtigosComprasDetalhe[Artigos]
5 | | | | | , "Regiao" , DISTINCTCOUNT(ZZPaíses[Regiao])
6 | | | | | , "Contar" , 1
7 | | | | | )
8
9 Var SKU_COUNT = SUMX(FILTER( Tabela_SKU, [Regiao] >1 ), [Contar])
10
11 return
12
13 SKU_COUNT
14

```

Figura 4.45- Medida "SKU from several regions"

Relativamente à medida da Figura 4.45, a variável “Tabela_SKU” cria uma tabela temporária utilizando a função *SUMMARIZE*, que agrupa os dados da tabela “zArtigosComprasDetalhe” com base na coluna de artigos. Para cada artigo, calcula-se o número de regiões distintas associadas a esse artigo, e é criada uma nova coluna chamada “Regiao” que contém esse número. A coluna “Contar” é simplesmente preenchida com o valor 1 para contar cada ocorrência.

A variável “SKU_COUNT” utiliza a função *SUMX* para iterar sobre a tabela “Tabela_SKU”. A função *FILTER* é usada para incluir apenas as linhas em que o número de regiões seja superior a 1, ou seja, apenas os artigos que estão presentes em mais de uma região. Para cada linha filtrada, a coluna “Contar” é somada, resultando no total de SKUs que estão presentes em várias regiões.

A medida “SKU from several regions” devolve o número de SKUs (artigos) que estão associados a mais de uma região. Isto é útil para entender a distribuição de produtos em diferentes regiões e monitorizar a presença geográfica dos mesmos.

As medidas DAX apresentadas anteriormente foram essenciais para a elaboração de relatórios interativos, que incluíam gráficos dinâmicos, tabelas de resumo e visuais personalizados, facilitando a compreensão dos resultados por parte dos utilizadores finais.

4.5.3 Desenvolvimento de relatórios

Os relatórios foram projetados para oferecer uma navegação intuitiva, onde os utilizadores podiam filtrar dados, explorar diferentes dimensões e obter indicadores relevantes de forma rápida e visualmente apelativa.

Durante o desenvolvimento destes relatórios, foram aplicadas boas práticas de desenvolvimento de relatórios e *storytelling* para garantir que os dados fossem comunicados de forma clara e impactante. Algumas das principais práticas incluíram:

Storytelling Aplicado: A técnica de *storytelling* foi implementada para orientar os utilizadores através dos dados, proporcionando uma narrativa clara e estruturada sobre as principais tendências e indicadores, facilitando a interpretação dos resultados e a sua aplicação no contexto empresarial. Na organização onde o estágio foi realizado, o *storytelling* dos relatórios era estruturado em três páginas principais, essenciais para a comunicação eficaz dos dados.

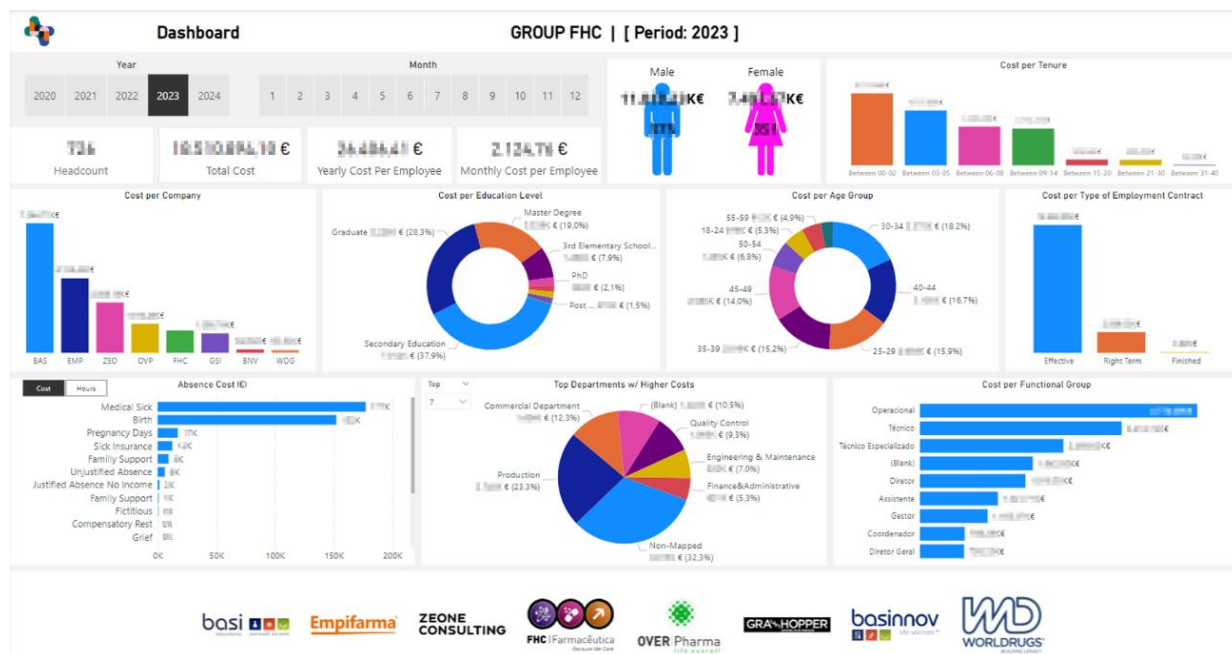


Figura 4.46- *Dashboard* do relatório "*People Cost*"

A primeira página, como demonstrado pela Figura 4.46, geralmente intitulada "*Dashboard*", apresentava uma visão geral dos principais indicadores e métricas-chave, oferecendo uma visão macro dos dados mais relevantes para a tomada de decisão. Esta

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

página funcionava como ponto de partida, destacando os indicadores mais críticos de uma forma visualmente apelativa e de fácil compreensão, permitindo que os utilizadores obtivessem uma visão imediata do estado dos processos ou resultados.

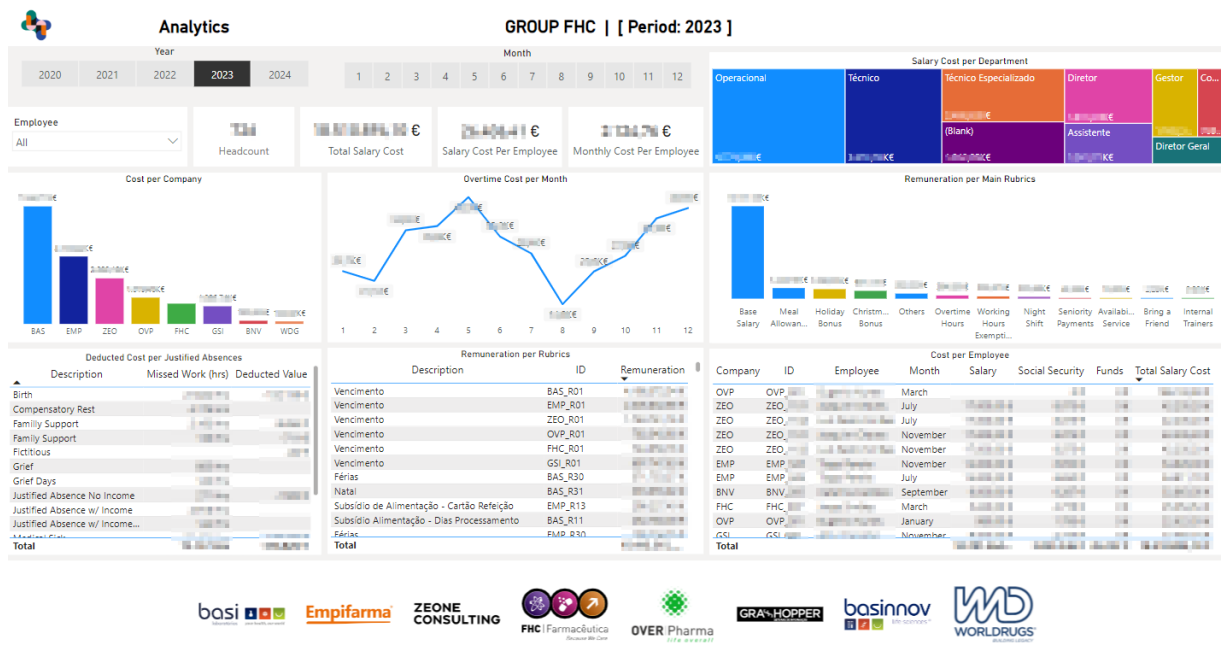


Figura 4.47- *Analytics* do relatório "People Cost"

A segunda página, com o título "*Analytics*", aprofundava a análise, combinando visualizações gráficas com dados mais detalhados, geralmente fornecidos por tabelas, como exibido na Figura 4.47. Esta página permitia uma exploração mais minuciosa dos dados, fornecendo contexto adicional e possibilitando a análise de relações e padrões mais complexos que não eram evidentes no *Dashboard*.

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

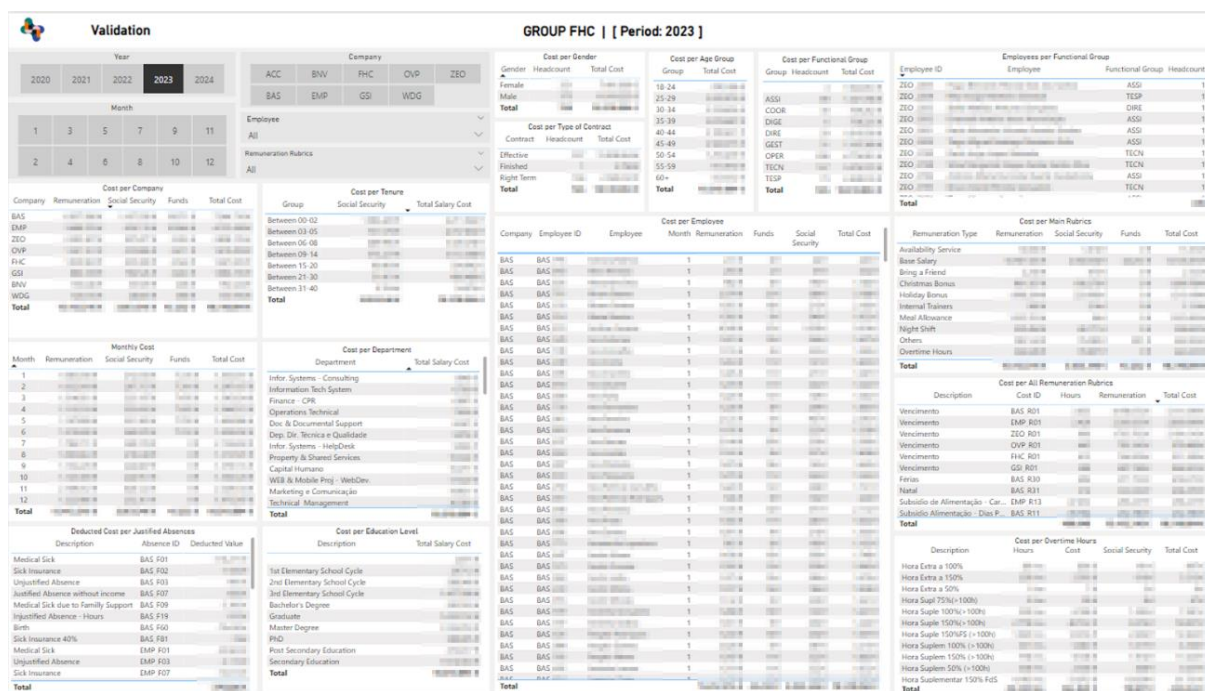


Figura 4.48- Validation do relatório "People Cost"

Por fim, a terceira página, denominada "Validation", era dedicada a filtros e tabelas que incluíam todos os dados apresentados nas páginas anteriores. Esta página permitia aos utilizadores aplicar filtros específicos para extrair detalhes adicionais dos dados, oferecendo um nível de personalização na análise que possibilitava a validação de informações ou o aprofundamento em áreas específicas de interesse. Através desta estrutura, o *storytelling* proporcionava uma navegação sequencial pelos dados, permitindo uma interpretação gradual, começando por uma visão geral até ao nível mais granular de detalhe, como apresentado pela Figura 4.48.

Estrutura e Clareza na Apresentação: A organização dos *dashboards* seguiu um layout lógico e intuitivo, facilitando a leitura e interpretação dos KPI principais. Na Figura 4.49, podemos observar um layout simples, minimalista e com um esquema de cores simples. Desta forma, os utilizadores conseguem retirar mais rapidamente as informações necessárias e não ficam sobrecarregados com visuais de difícil interpretação.

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

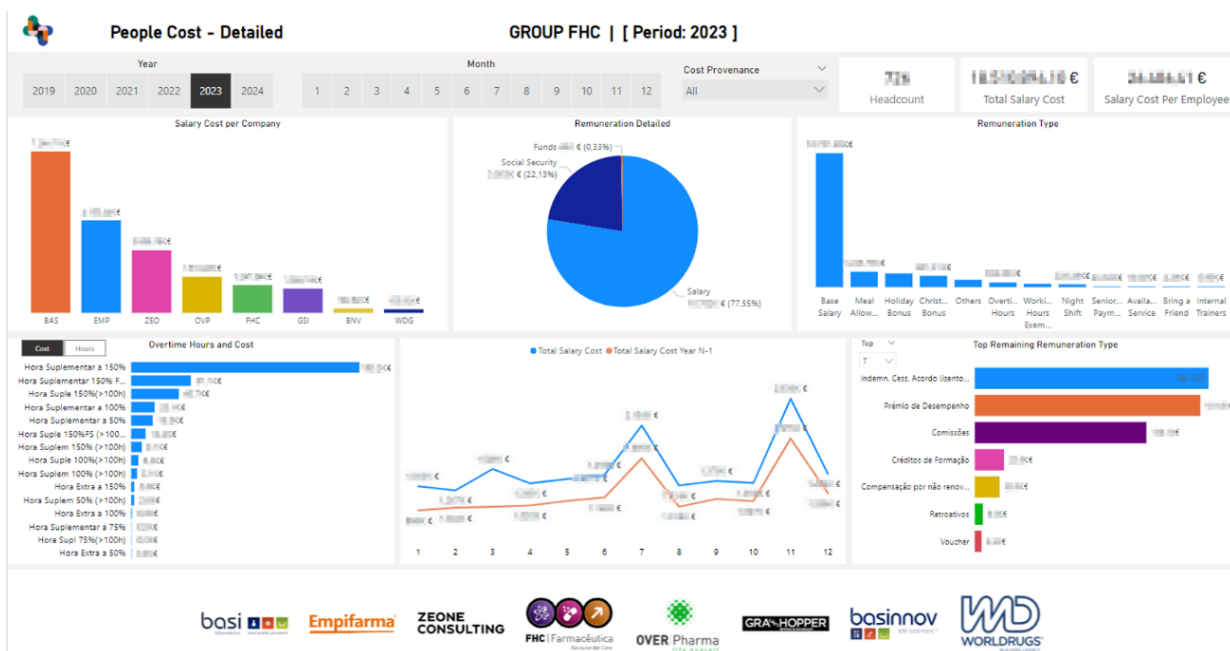


Figura 4.49- *Detailed* do relatório "People Cost"

Utilização de *Dashboards* Interativos: Os relatórios permitiam que os utilizadores explorassem os dados de forma interativa, disponibilizando filtros e aplicando *tooltips* para visualizar detalhes mais específicos. Na Figura 4.50, observamos uma *dashboard* com disponibilização de filtros para que o utilizador conseguisse retirar conclusões mais pormenorizada. Como apresentado na Figura 4.51, o recurso a *tooltips* foi utilizado para fornecer informações mais específicas. Assim, o utilizador conseguia ter acesso a uma maior quantidade de informações, sem ser sobrecarregado com um número excessivo de visuais.

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

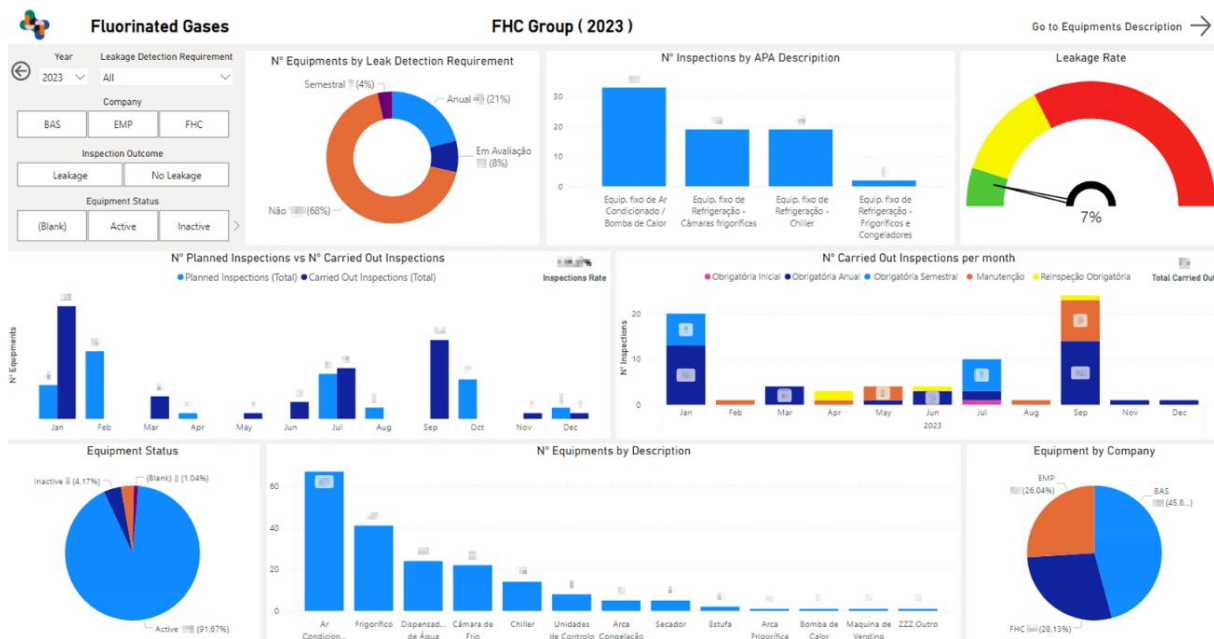


Figura 4.50- Dashboard com Slicers do relatório "Fluorinated Gases"

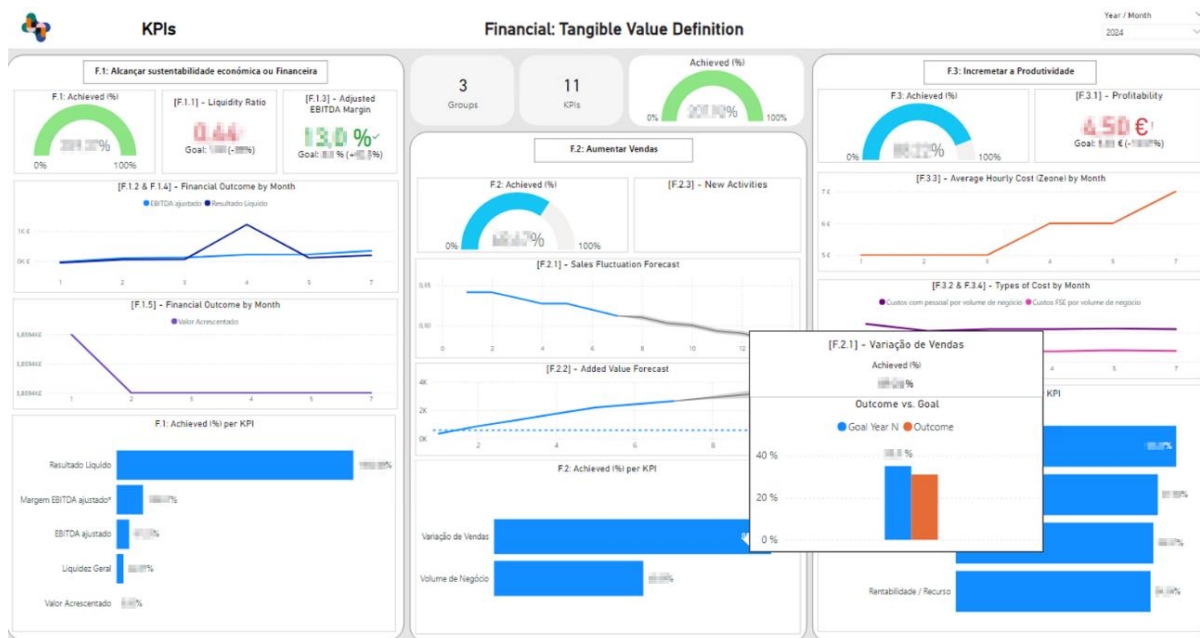


Figura 4.51- KPI Financeiros com Tooltip no relatório "Strategic Planning"

Filtragem de Dados Irrelevantes: Apenas os dados relevantes foram apresentados nos relatórios, assegurando a sua eficiência e evitando sobrecargas. Como demonstrado pela Figura 4.52, foi possível aplicar filtros diretamente na página para exibir apenas os dados

considerados pertinentes. Em certos casos, os dados em branco não são filtrados intencionalmente, com o objetivo de mostrar ao utilizador quais são os dados que ainda não foram mapeados. Na Figura 4.53, pode-se observar que filtrar esses dados da página resultaria numa apresentação incorreta do total do custo, o que poderia induzir o utilizador em erro.

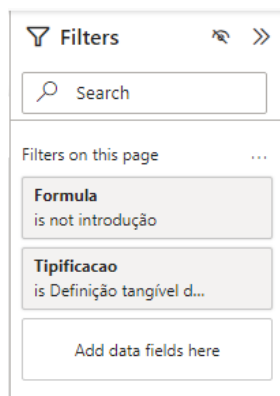


Figura 4.52- Painel de filtros aplicados diretamente na página

| Group | Headcount | Total Cost |
|--------------|-----------|------------|
| ASSI | | |
| COORD | | |
| DIGE | | |
| DIRE | | |
| GEST | | |
| OPER | | |
| TECN | | |
| TESP | | |
| Total | | |

Figura 4.53- Valores em branco no relatório "*People Cost*"

Design Eficaz e Consistência: O design dos relatórios seguiu princípios de clareza, com o uso estratégico de cores para destacar informações importantes e a manutenção de uma consistência visual para melhorar a experiência do utilizador. Como observado na Figura 4.54, a utilização da formatação condicional da cor da letra nos cartões de variação teve como objetivo facilitar a interpretação das informações pelos utilizadores. Ao alterar automaticamente a cor da letra com base nas variações dos indicadores, os utilizadores conseguiram identificar de forma imediata se uma variação é positiva ou negativa. Nas

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

tabelas, como demonstrado pela Figura 4.55, a formatação condicional foi aplicada através da cor de fundo das células em colunas específicas, permitindo destacar visualmente os valores mais relevantes. Estes recursos visuais tornaram os dados mais acessíveis e permitiram uma tomada de decisão mais rápida e informada.

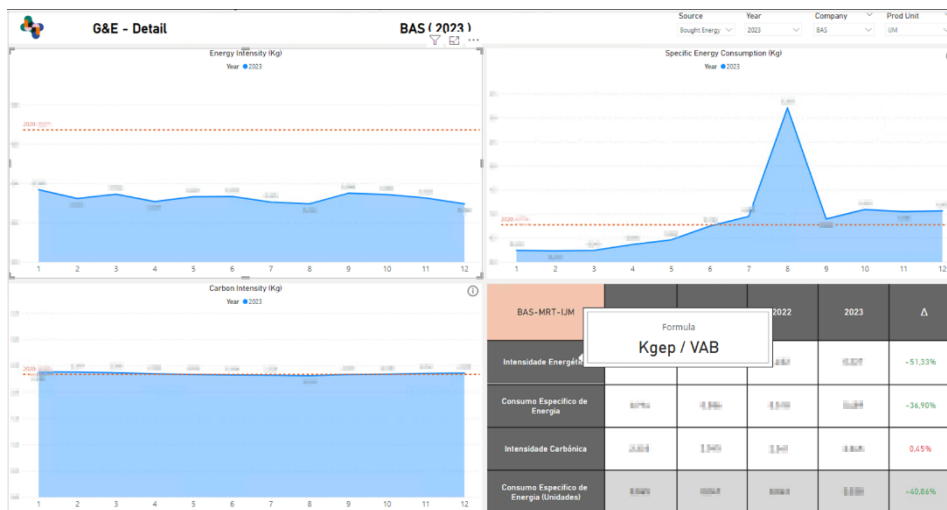


Figura 4.54- Formatação de cores e *Tooltip* com fórmulas no relatório "Utilities"

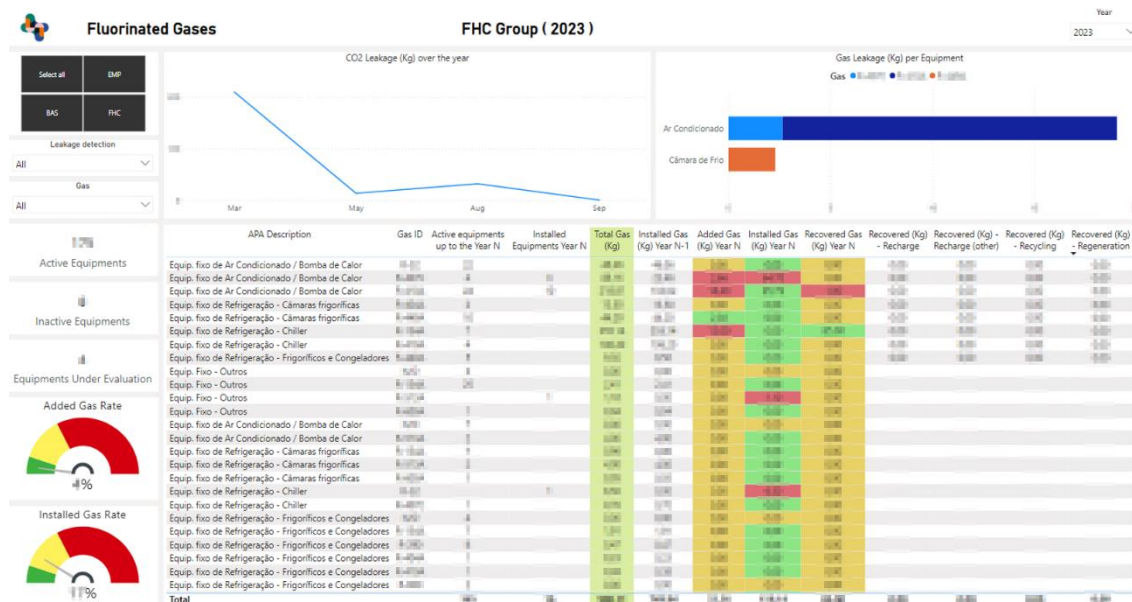


Figura 4.55- Formatação de cores das células da tabela no relatório "Fluorinated Gases"

A criação destes modelos ER e a integração das medidas DAX proporcionaram uma estrutura robusta para o relatório de análise de dados. As visualizações e gráficos facilitaram a interpretação dos resultados, permitindo que os utilizadores explorassem os dados com diferentes filtros e dimensões, ajudando-os a tomar decisões informadas com base em dados precisos e atualizados.

4.6 Avaliação dos modelos

A fase de avaliação de modelos foi uma das mais importantes no processo CRISP-DM, pois é nesta fase que se verificou a precisão e a adequação dos modelos construídos. O objetivo desta fase foi garantir que os dados gerados pelos modelos semânticos fossem reais e consistentes, de modo a fornecer indicadores corretos aos utilizadores.

Durante esta fase, foi essencial filtrar os dados nos relatórios para validar se os dados do modelo refletiam corretamente a realidade do negócio. A verificação envolveu a aplicação de filtros específicos nos *dashboards* e relatórios para identificar possíveis discrepâncias ou erros nos dados. Isso permitiu avaliar se os KPI e métricas fornecidos eram coerentes e utilizáveis para a tomada de decisão.

Em alguns casos, foi necessário criar medidas em DAX com o objetivo de validar a veracidade dos dados, confirmando que os cálculos e métricas refletiam corretamente os eventos ou transações capturadas pelos modelos de dados. As medidas DAX funcionaram como uma forma de garantir que as somas, médias, percentagens e outras operações calculadas estivessem alinhadas com as expectativas e as necessidades do negócio.



Figura 4.56- Medida de validação de número artigos no relatório "*Chemical Products*"

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

Como observado na Figura 4.56, a medida criada inicialmente não estava a fornecer os valores corretos, o que indicou uma discrepância nos resultados apresentados. Dado que a precisão dos dados é fundamental para garantir a confiabilidade do relatório e a correta interpretação dos indicadores, foi necessário ajustar a medida em DAX para obter os valores reais.

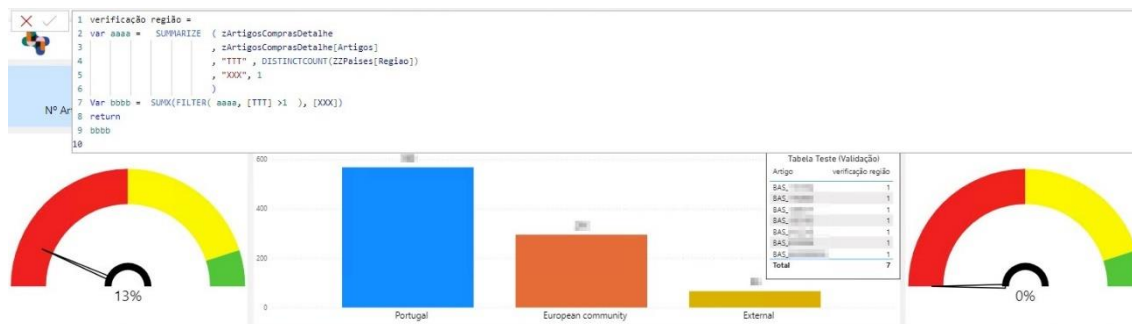


Figura 4.57- Medida de validação após correção no relatório "Chemical Products"

A Figura 4.57 apresenta a correção da medida envolveu a revisão das expressões DAX aplicadas, verificando as fórmulas utilizadas e ajustando os cálculos para garantir que os valores refletissem corretamente os dados subjacentes. Este ajuste foi importante para assegurar que o gráfico apresentasse indicadores precisos, permitindo aos utilizadores tomar decisões com base em dados corretos.

| IDLinha | Doc | Serie | N.º | Data | Entidade | Pais | Descrição | Unidade | Quantidade | Preço Líquido | Total Líquido | Artigos | ArtigoID | statecodename | statuscodename | _Quantidade KILO | Qtd. KILOS | Cov. Quantidade % | ADR |
|---------|----------------|-------|-----|------|----------|------|-----------|---------|------------|---------------|---------------|---------|-------------|---------------|----------------|------------------|------------|-------------------|-----|
| 1612 | EE593952-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1613 | EE593951-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1614 | EE59394C-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1615 | EE59394B-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1616 | EE59394A-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1617 | EE593948-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1618 | EE593946-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1619 | EE59393E-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1620 | EE59393B-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1621 | EE59393A-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1622 | EE593935-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1623 | EE593931-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1624 | EE59392E-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1625 | EE593929-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1626 | EE593928-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1627 | EE593920-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1628 | EE59391F-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1629 | EE59391D-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1630 | EE593919-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1631 | EE593918-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1632 | EE593915-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1633 | EE593914-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1634 | EE593913-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1635 | EE59390F-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1636 | EE59390C-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1637 | EE593905-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1638 | EE593902-J-VFA | | | | | PT | | | | | | | C4DB601A-J | | | | | | |
| 1639 | 326R2A96-J-VFA | | | | | PT | | | | | | | C3A42C50-J | | | | | | |
| 1640 | 326B2A95-J-VFA | | | | | PT | | | | | | | C3A42C50-J | | | | | | |
| 1641 | B91D2856-J-VFA | | | | | PT | | | | | | | E2851BA3-J | | | | | | |
| 1642 | 482C33FB-J-VFA | | | | | PT | | | | | | | E8E37048-A | | | | | | |
| 1643 | 482C33FA-J-VFA | | | | | PT | | | | | | | C7E4D264C-J | | | | | | |
| 1644 | 48636599-J-VFA | | | | | PT | | | | | | | E2851BA3-J | | | | | | |
| 1645 | 37AD6431-J-VFA | | | | | PT | | | | | | | CLF42D45-J | | | | | | |
| 1646 | BDE48F44-J-VFA | | | | | PT | | | | | | | C02F7842-A | | | | | | |
| 1647 | | | | | | | | | | | | | | | | | | | |
| 1648 | | | | | | | | | | | | | | | | | | | |

Figura 4.58- Tabela dinâmica para validação da coluna "_Quantidade KILO"

Noutros casos, como exemplificado na Figura 4.58, a validação foi realizada através da exportação dos dados para o *Excel*, onde foram criadas tabelas dinâmicas que permitiram uma verificação manual da consistência e precisão dos dados.

| Nº ONU (1) | Artigo | Total Quantidade (Kg) |
|--------------|--------|-----------------------|
| UN 3261 | BAS_ | |
| UN 2811 | BAS_ | |
| UN 3077 | BAS_ | |
| UN 2811 | BAS_ | |
| UN 3077 | BAS_ | |
| UN 3077 | BAS_ | |
| UN 2811 | BAS_ | |
| UN 2810 | BAS_ | |
| UN 2811 | BAS_ | |
| UN 2811 | BAS_ | |
| UN 3077 | BAS_ | |
| UN 3077 | BAS_ | |
| UN 2811 | BAS_ | |
| UN 2811 | BAS_ | |
| UN 2811 | BAS_ | |
| UN 2811 | BAS_ | |
| UN 3077 | BAS_ | |
| Total | | 131.438,64 |

Figura 4.59- Verificação do resultado da medida criada

Através da Figura 4.59 é possível concluir que a medida criada estava em conformidade e apresentava dados verídicos. Esta abordagem foi particularmente útil para realizar comparações detalhadas e garantir que os valores exibidos nos relatórios correspondiam aos dados subjacentes, proporcionando uma camada extra de verificação.

Esta fase foi crucial para assegurar que o modelo não só estivesse bem estruturado, mas também que as análises fossem precisas e confiáveis. A validação contínua dos modelos foi fundamental para garantir a confiança nos dados apresentados, assegurando que os utilizadores pudessem tomar decisões baseadas em dados reais e fiáveis.

4.7 Implantação

Na última fase do CRISP-DM, o foco esteve na implantação e publicação dos modelos semânticos resultantes das fases anteriores. Nesta etapa, os resultados das análises foram transformados em ações concretas, com o objetivo de melhorar os processos em diversas áreas da empresa e melhorar a eficiência operacional. A configuração do RLS foi

implementada de acordo com as permissões de visualização definidas internamente. Dado que os relatórios eram disponibilizados a diferentes audiências dentro da organização, como observado na Figura 4.60, esta configuração assegurava que os utilizadores não acediam a dados relativos a outras empresas do grupo, promovendo um controlo rigoroso e seguro sobre as informações visualizadas por cada grupo

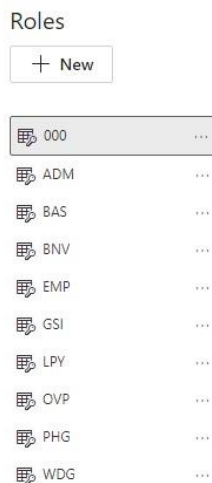


Figura 4.60 - Security Roles

Posteriormente, aquando da publicação dos relatórios no *Power BI Service*, eram definidas também audiências para restringir as visualizações dos utilizadores somente ao necessário de cada perfil. Através desta disponibilização, os utilizadores passaram a ter acesso a relatórios dinâmicos que suportavam a tomada de decisões informadas.

Esta fase foi essencial para garantir que as análises não fornecessem apenas indicadores, mas também se traduzissem em melhorias operacionais concretas, contribuindo para a melhoria contínua na gestão das operações. Após a publicação dos modelos semânticos e dos relatórios correspondentes, foram-se ajustando os modelos semânticos de acordo com os objetivos da empresa e o *feedback* dos utilizadores.

Foi também nesta fase que os utilizadores reportaram situações em que era necessário verificar o modelo construído. Por exemplo, durante a implantação, foram observadas algumas falhas nos dados que estavam a ser apresentados.

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

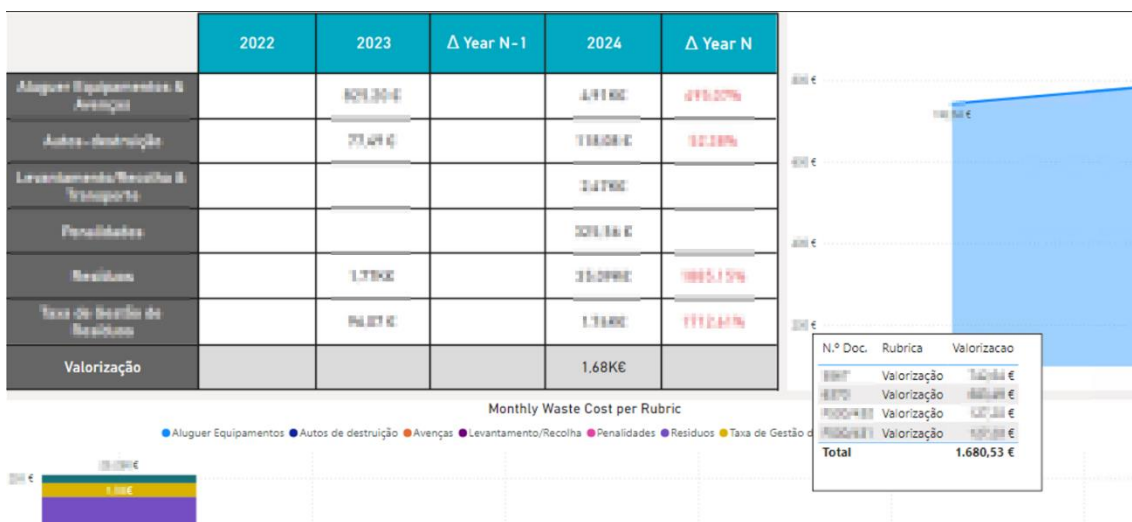


Figura 4.61- Verificação de documentos de valorização em falta

Como se observa na Figura 4.61, haviam dados que não estavam a aparecer corretamente no relatório. Para diagnosticar o problema, foi necessário criar uma tabela de verificação para identificar quais faturas estavam em falta.

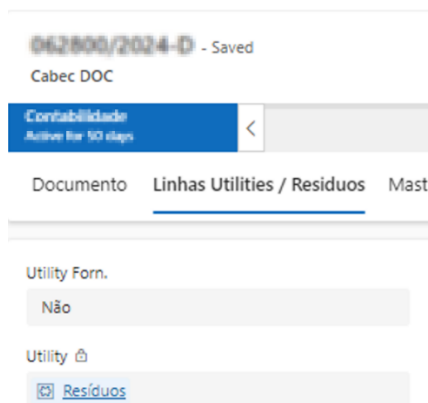


Figura 4.62- Introdução de dados incorretos no campo "Utility Forn."

Na Figura 4.62, foi identificado que o problema se devia a uma falha na introdução dos dados no *Dataverse*, causando discrepâncias nos dados apresentados.

Desenvolvimento de Modelos Semânticos: *Business Intelligence* na Zeone Consulting

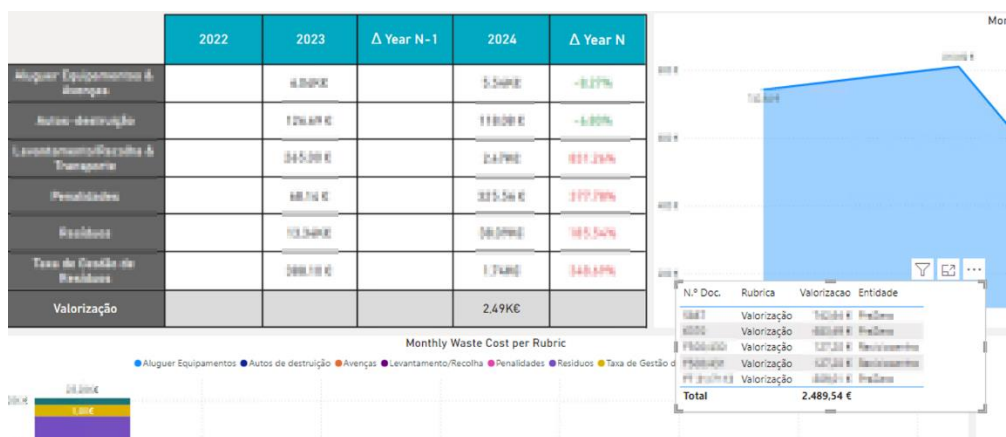


Figura 4.63- Verificação de documentos após atualização dos dados

Após a correção, conforme demonstrado na Figura 4.63, os dados foram inseridos corretamente no *Dataverse* e passaram a ser exibidos corretamente nos relatórios, confirmando que o modelo estava a funcionar de forma adequada.

Esta fase garantiu que as análises e modelos semânticos fossem implementados com sucesso, assegurando que os dados fossem constantemente atualizados e ajustados, conforme as necessidades e o *feedback* dos utilizadores, promovendo a continuidade e a eficácia operacional.

CONCLUSÃO

Ao longo deste estágio na *Zeone Consulting*, empresa pertencente ao Grupo FHC, os trabalhos realizados tiveram especial foco no uso do *software Power BI*. A metodologia do trabalho foi adaptada da metodologia CRISP-DM, organizando as atividades desde a compreensão dos dados e do negócio até ao desenvolvimento de modelos semânticos e implantação dos mesmos.

A motivação principal aquando da realização do estágio foi contribuir para a melhoria dos processos de análise de dados da empresa no desenvolvimento dos modelos semânticos. A combinação da garantia da qualidade e integridade dos dados e a aplicação de boas práticas de desenvolvimento de relatórios permitiu a tomada de decisões mais assertivas.

Os principais objetivos foram cumpridos, incluindo a criação de modelos ER e a aplicação de boas práticas de desenvolvimento de relatórios no *Power BI*, resultando em visualizações interativas e claras. A criação de medidas em DAX permitiu a elaboração de relatórios que suportaram a tomada de decisões com base em dados precisos e fiáveis, garantindo assim a qualidade e integridade das análises. A implementação do *storytelling* na comunicação dos resultados foi um contributo significativo para melhorar a compreensão dos dados por parte dos utilizadores.

Uma das principais dificuldades encontradas foi garantir que os dados obtidos através das métricas fossem sempre verídicos. Esse desafio exigiu uma validação constante das medidas e uma verificação minuciosa das fontes de dados, especialmente quando surgiam discrepâncias entre os relatórios e os dados reais. Este processo foi fundamental para assegurar a confiança nos dados apresentados.

No que toca a limitações, além do tempo extra necessário para validar os dados e garantir a veracidade das métricas, outro ponto a destacar foi a complexidade na integração e preparação de dados provenientes de múltiplas fontes, como *SharePoint*, *Dataverse* e *SQL*, cada uma com os seus desafios específicos.

Para trabalhos futuros, sugere-se a exploração de técnicas mais avançadas de análise preditiva para melhorar a análise dos dados. Também seria útil continuar a automatizar a validação dos dados, minimizando o esforço manual e garantindo a precisão das métricas.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alasiri, M. M., & Salameh, A. A. (2020). The impact of business intelligence (BI) and decision support systems (DSS): Exploratory study. *International Journal of Management*, 11(5), 1001–1016. <https://doi.org/10.34218/IJM.11.5.2020.092>
- Bansal, A. (2023). Power BI Semantic Models to enhance Data Analytics and Decision-Making. *International Journal of Management*, 14(5), 136–142. https://iaeme-library.com/index.php/IJM/article/view/IJM_14_05_012
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step Data Mining Guide* (SPSS Incorporated, Ed.; pp. 1–76). CRISP-DM Consortium.
- Dhanoa, V., Hinterreiter, A., Fediuk, V., Elmqvist, N., Groeller, E., & Streit, M. (2024). *StoryStore: Interactive Storytelling for Dashboard Onboarding* (Vol. 43, Issue 3).
- Group FHC (2024). Um Grupo global | Group FHC – The Future of Health care. <https://www.fhcthefutureofhealthcare.pt/pt/um-grupo-global/> (consultado a 16/09/2024)
- Heinrich, B., Hristova, D., Klier, M., Schiller, A., & Szubartowicz, M. (2017). Requirements for data quality metrics. *Journal of Data and Information Quality*, 9(2), 1–32. <https://doi.org/10.1145/3148238>
- Kaushik, A. (2024). *Explain the concept of Report View, Data View, and Model View in Power BI*. Tutorialspoint. <https://www.tutorialspoint.com/explain-the-concept-of-report-view-data-view-and-model-view-in-power-bi> (consultado a 16/09/2024)
- Laudon, K. C., & Laudon, J. P. (2022). *Management Information Systems: Managing the Digital Firm, 17th Edition*. Pearson.
- Martins, N., Martins, S., & Brandão, D. (2022). *Design Principles in the Development of Dashboards for Business Management*. In *Springer Series in Design and Innovation* (Vol. 16, pp. 353–365). Springer Nature. https://doi.org/10.1007/978-3-030-79879-6_26

- Microsoft Learn. (2023). *Understand star schema and the importance for Power BI - Power BI*. Microsoft Learn. <https://learn.microsoft.com/en-us/power-bi/guidance/star-schema> (consultado a 16/09/2024)
- Microsoft Learn. (2024). *What is Power BI? - Power BI*. Microsoft Learn. <https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview> (consultado a 28/08/2024)
- Microsoft Learn. (2024). *What is Power Query? - Power Query*. Microsoft Learn. <https://learn.microsoft.com/en-us/power-query/power-query-what-is-power-query> (consultado a 28/08/2024)
- Mishra, B. K., Hazra, D., Tarannum, K., & Kumar, M. (2017). *Business Intelligence using Data Mining techniques and Business Analytics. Proceedings of the 5th International Conference on System Modeling and Advancement in Research Trends, SMART 2016*, 84–89. <https://doi.org/10.1109/SYSMART.2016.7894496>
- Numerro. (2022). *The complete guide to designing Power BI reports*. Numerro. <https://www.numerro.io/guides/the-complete-guide-to-designing-power-bi-reports> (consultado a 16/09/2024)
- Cotton, R. (2022). *DAX Cheat Sheet*. Datacamp. <https://www.datacamp.com/cheat-sheet/dax-cheat-sheet> (consultado a 20/09/2024)
- Romero, D., Flores, M., Herrera, M., & Resendez, H. (2019, June 1). Five Management Pillars for Digital Transformation Integrating the Lean Thinking Philosophy. *Proceedings - 2019 IEEE International Conference on Engineering, Technology and Innovation, ICE/ITMC 2019*. <https://doi.org/10.1109/ICE.2019.8792650>
- Rouhani, S., Ashrafi, A., Zare Ravasan, A., & Afshari, S. (2016). The impact model of business intelligence on decision support and organizational benefits. *Journal of Enterprise Information Management*, 29(1), 19–50. <https://doi.org/10.1108/JEIM-12-2014-0126>
- Rubin, V. L. (2014). Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online*, 24, 4–15. <https://doi.org/10.7152/acro.v24i1.14671>

Singh, T. (2023). 10 Business Intelligence Best Practices for 2024. ScaleupAlly. <https://scaleupally.io/blog/business-intelligence-best-practices/> (consultado a 10/09/2024)

Tvrđíková, M. (2007). *Support of Decision Making by Business Intelligence Tools*. In Proceedings – 6th International Conference on Computer Information Systems and Industrial Management Applications, CISIM 2007 (pp.364-368). IEEE Computer Society.

Vikash, N., & Kumar, T. V. V. (2023). Veracity assessment of big data. In Lecture notes in networks and systems (pp. 305–315). https://doi.org/10.1007/978-981-99-5088-1_26

Wexler, S., Shaffer, J., & Cotgreave, A. (2017). *The big book of dashboards: visualizing your data using real-world business scenarios*. Wiley.

Zeone Consulting. (2024). Zeone Consulting. <https://www.zeone.pt/> (consultado a 10/09/2024)