

1 2



9 0

FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE DE  
COIMBRA

## **Inteligência no Negócio 2023-24**

### **Consumo de energia nos Países Baixos**

**Professora: Catarina Helena Branco Simões da Silva**

#### **Grupo 1**

**Diogo Alexandre Santos Rosário nº 2023185395**

**João Eduardo Monteiro Raposo nº 2023147060**

## ÍNDICE

<b>1. INTRODUÇÃO.....</b>	<b>3</b>
<b>2. CONTEXTO.....</b>	<b>4</b>
<b>3. PROPOSTA DE VALOR.....</b>	<b>5</b>
<b>4. DATA.....</b>	<b>6</b>
<b>5. TECNOLOGIA DE DW/BD.....</b>	<b>7</b>
<b>6. MODELO ESTRELA.....</b>	<b>8</b>
<b>7. ETL.....</b>	<b>9</b>
<b>8. OLAP.....</b>	<b>11</b>
<b>9. ESQUEMAS E MOCKUPS DO POWER BI.....</b>	<b>12</b>
<b>10. CONCLUSÃO.....</b>	<b>15</b>

# 1. INTRODUÇÃO

Com o presente documento, o grupo tem como objetivo apresentar uma solução alinhada com os princípios fundamentais de *Business Intelligence (BI)*. Esta proposta de valor aborda diversos elementos essenciais, desde da ideia de negócio até à implementação técnica, destacando a relevância da análise dos dados recolhidos com o intuito de impulsionar a tomada de decisões empresariais inteligentes e eficazes.

Ao longo deste relatório, vão ser explorados os componentes cruciais que compõem a nossa abordagem de *Business Intelligence*. Inicialmente, será delineada a ideia de negócio subjacente à solução proposta, destacando as suas características e o seu potencial impacto no contexto do problema. Em seguida, serão detalhados ao pormenor os processos de recolha e tratamento de dados, dada importância também às estratégias adotadas para garantir a qualidade e a relevância das informações obtidas.

Além disso, serão apresentados *mockups* da possível solução desenvolvida, proporcionando uma visão concreta e tangível do nosso sistema. Estes ilustram uma ideia da interface e as funcionalidades chave da solução, permitindo uma compreensão simples e clara de como os nossos *stakeholders* poderão interagir com o sistema.

No que diz respeito às tecnologias utilizadas, serão discutidas as escolhas fundamentais relacionadas à arquitetura de dados, incluindo a tecnologia de *Data Warehouse* e a base de dados utilizada. Serão abordados também os processos de extração, transformação e carga (*ETL*), que desempenham um papel crucial na integração e preparação de dados para análise.

Por fim, serão exploradas as soluções de *Online Analytical Processing (OLAP)*, destacando a sua relevância na análise e visualização dos dados de forma dinâmica e interativa. Estes aspetos combinados formam uma base sólida para a aplicação eficaz de BI, dando as organizações a possibilidade de extrair *insights* valiosas e a tomar decisões fundamentais com base em evidências sólidas

## 2. CONTEXTO

No contexto atual, a energia desempenha um papel super crucial para o funcionamento de todas as facetas da nossa sociedade. À medida que avançamos em caminho do futuro, torna-se cada vez mais evidente que a energia será a peça principal sobre qual toda a estrutura sociedade se baseia e opera. Neste sentido, é essencial reconhecer a importância vital da distribuição eficiente de energia, especialmente à luz do aumento do consumo decorrente da digitalização do nosso mundo real.

Com a evolução da sociedade em direção a esta digitalização e ao surgimento da inteligência artificial, a demanda de energia está a experimentar um crescimento significativo, impulsionado pelo aumento do número de dispositivos conectados e pela expansão de tecnologias emergentes.

O aumento do uso de dispositivos inteligentes, sistemas de automação e tecnologias de comunicação consome energia de forma contínua, exigindo uma abordagem inovadora na distribuição e gestão deste recurso.

Portanto, torna-se evidente que a integração de soluções inteligentes é essencial para enfrentar os desafios emergentes relacionados à distribuição de energia. Ao compreender profundamente os padrões de consumo e as necessidades energéticas em constante evolução, as empresas que distribuem energia podem adaptar e/ou melhorar as suas estratégias tornando-as mais eficientes e personalizadas.

### 3. PROPOSTA DE VALOR

#### 3.1 Quem é o nosso cliente (*Stakeholder*)?

Qualquer empresa que efetue a gestão da rede elétrica e distribuição de energia num país ou cidade, que pretenda estudar a concorrência, analisar as suas métricas ao final do ano e fazer previsões da quantidade de energia que irá ser consumida nos próximos anos.

#### 3.2 Proposta de valor

Propomos a criação de um sistema que torne possível acompanhar as tendências de consumo ao longo do tempo e identificar padrões de consumo energético em diferentes regiões. As análises podem abranger desde a disseminação de contadores inteligentes até a produção de energia doméstica, como painéis solares e até mesmo a influência de veículos elétricos no consumo energético local.

Com uma fonte confiável e abrangente de dados, empresas e entidades do setor energético podem tomar decisões informadas e estratégicas, antecipando-se às necessidades futuras e contribuindo para uma gestão mais eficiente e sustentável da energia no país.

#### 3.3 Indicadores-Chave de desempenho do negócio.

1. **Taxa de crescimento de contadores inteligentes:** Taxa onde é medida a velocidade com que os contadores inteligentes estão a ser adotados, refletindo assim na modernização da infraestrutura e a capacidade de monitorizar o consumo de energia de forma mais eficiente.
2. **Variação no consumo de energia residencial:** Acompanha as tendências de consumo de energia em residências ao longo do tempo, podendo ajudar a identificar padrões sazonais, influência de fatores externos (como mudanças climáticas) e eficácia de campanhas de eficiência energética.
3. **Adoção de energias renováveis (por exemplo, solar):** Avalia o crescimento da produção de energia renovável, como a energia solar, indicando a transição para fontes mais sustentáveis de energia.
4. **Previsão do consumo de energia no futuro:** Utiliza dados históricos para prever o consumo de energia no próximo ano, sendo crucial para o planeamento do fornecimento adequado da mesma.

#### 3.4 Principais desafios que a empresa enfrenta para crescer

Os principais desafios que a empresa enfrenta para crescer são variados e podem incluir:

1. **Inovação Tecnológica:** Garantir que a empresa esteja sempre atualizada com as últimas tecnologias e tendências do setor energético, como a implementação de medidores inteligentes e soluções de energia renovável, para oferecer serviços mais eficientes e sustentáveis aos clientes.
2. **Competitividade:** É essencial manter-se competitivo e procurar oportunidades de diferenciação em termos de serviços, eficiência operacional e satisfação do cliente para se destacar entre as outras empresas e atender às expectativas crescentes dos consumidores.
3. **Gestão de Dados:** Lidar de forma eficiente com grandes volumes de dados, garantindo a sua qualidade, segurança e utilização adequada para análises e tomadas de decisão estratégicas.

## 4. DATA

Após uma pesquisa cuidadosa, os dados foram obtidos da plataforma **Kaggle**, cuja é conhecida por conter uma vasta variedade de *datasets* públicos usados em projetos de *Data Science* e *Machine Learning*. Desta plataforma, foi retirado um conjunto de dados (*dataset*) que contém uma variedade de informações de diversas empresas de distribuição de energia dos Países Baixos.

O *Dataset* contém uma série de informações recolhidas por várias distribuidoras de energia nos Países Baixos nomeadamente: **Enexis**, **Liander**, **Stedin**, **Enduris**, **Westlandinfrac**, **Rendo** e **Coteq**. Todos os dados estão normalizados e contém 14 *características* que são: o código do administrador da rede, a área onde a energia está a ser consumida, o nome da rua, a lista de códigos postais, a cidade, o número de conexões, a percentagem de eletricidade consumida, a percentagem de conexões ativas, o consumo anual de eletricidade, o consumo anual de eletricidade nas horas de tarifário mais baixo e a percentagem de contadores inteligentes. Além destas informações, os dados estão disponíveis em formato CSV, desde 2011 até 2020 e cada ficheiro corresponde a consumos anuais de cada ano. É importante também lembrar que o *dataset* tem exatamente 146 *megabytes* e os dados, desde 2020 que não são atualizados na plataforma.

Para além dos dados sobre a energia distribuída nos Países Baixos, recolhemos também um outro *dataset* sobre a temperatura e as suas variáveis associadas ao longo dos anos no mesmo país em questão. Os dados são desde 1901 até 2020 e são dados recolhidos ao longo dos dias de vários anos.

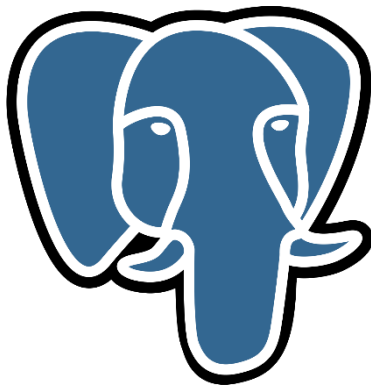
Para finalizar, foi encontrado nos dados alguns problemas relativos a: colunas adicionais em certos ficheiros que indicavam informação desnecessária sobre o caso de estudo, existência de várias ocorrências de nomes parecidos para a mesma empresa em questão, colunas e valores trocados por outras. Para todos estes problemas, o grupo aplicou técnicas de seleção e transformação de dados para corrigir estes problemas.

## 5. TECNOLOGIA DE DW / BD

As tecnologias utilizadas para criar e testar a solução mencionada nos capítulos anteriores foram diversas, mas destacamos essencialmente o uso do *Pentaho* para selecionar e carregar os dados selecionados para a nossa base de dados *PostgreSQL*.

O *Pentaho* é uma plataforma de integração de dados que oferece um conjunto de ferramentas que são úteis para transformar, limpar e carregar dados de várias extensões diferentes. Esta ferramenta permitiu-nos automatizar o processo da integração dos dados recolhidos provenientes da plataforma *Kaggle* de forma eficiente, garantindo que estivessem prontos para serem armazenados na nossa base de dados.

Adicionalmente, o PostgreSQL é um sistema que dá a possibilidade ao utilizador gerir bases de dados relacionais. Também é conhecido por ser um sistema robusto, escalável e altamente personalizável, sendo amplamente utilizado em aplicações de várias empresas e/ou associações devido à sua capacidade de lidar com grandes volumes de dados.



*Figura 1 - PostgreSQL*



*Figura 2 - Pentaho*

## 6. MODELO ESTRELA

A figura 3 representa o esquema em estrela para este projeto de inteligência no negócio. As principais características do esquema são:

### Tabela de Factos *energy*:

Na tabela de fatos *energy*, é possível observar que será aqui feita a apresentação dos tipos de energia, como a energia consumida anualmente e a percentagem de energia consumida de tarifa baixa e também a apresentação de energia devolvida à rede. Estas três variáveis serão os nossos factos. As restantes variáveis, são os ID's das dimensões construídas e usadas ao longo do projeto: ID da linha de distribuição, o ID da meteorologia e o tempo, neste caso em anos.

### Tabelas de Dimensões:

Para as tabelas de dimensões, são apresentadas três:

***distribution\_line*** – Nesta tabela estão contidos os atributos que dão informações referentes às linhas de distribuição de energia das empresas que informam por exemplo, de quantas conexões uma certa linha contém, em que cidades e ruas estão distribuídas, quantos contentores inteligentes existem, etc.

***weather*** – Nesta tabela estão contidos os atributos que dão informações referentes às condições meteorológicas nos Países Baixos. Os atributos poderão ser observados através do segundo link da bibliografia. Alguns representam os níveis percentuais de humidade ou de quantidade de horas que choveu em certos anos.

***time*** – Esta tabela é usada aqui para representar que medida de tempo será usado no projeto. A medida de tempo a usar foi em anos, pois os dados acerca do consumo energético nos Países Baixos estavam expressos em anos.

**Relações:** As relações apresentadas são M:1, onde a uma energia poderá pertencer a muitas linhas de distribuição, bem como poderão ser representadas várias energias num dado único meteorológico ou num ano.

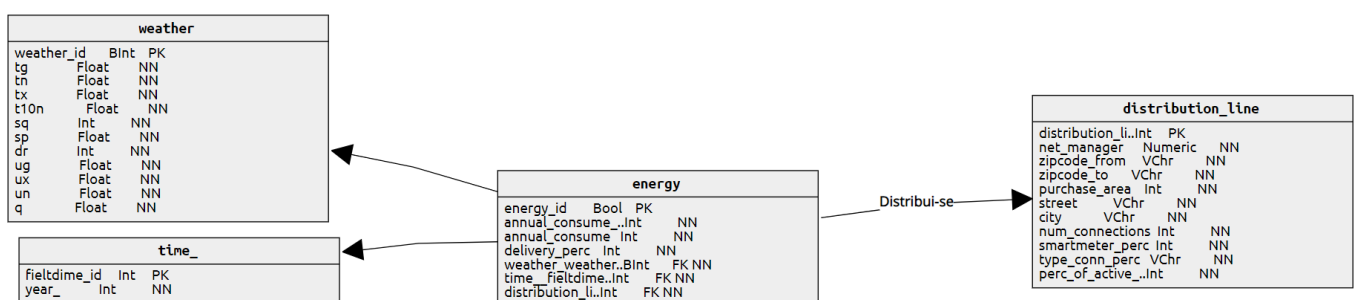


Figura 3 – Esquema Estrela do projeto.



## 7. ETL

Neste capítulo, pretendemos detalhar todo o processo de *Extract, Transform and Load (ETL)* que seguimos, bem como todo o planeamento associado a estas etapas.

Inicialmente, pusemos em prática o processo de extração de dados recolhidos, conforme foi descrito nos capítulos anteriores. Além dos dados mencionados, decidimos também incorporar informações meteorológicas dos Países Baixos, abrangendo um período extenso desde 1901 até 2020. Esta decisão estratégica permitiu-nos investigar possíveis correlações entre o consumo de energia e as condições climáticas, com o intuito de identificar possíveis padrões que pudessem contribuir para uma análise mais precisa.

Num segundo momento, dedicámo-nos à etapa de transformação dos *datasets*. Neste processo, procedemos a uma limpeza dos dados inconsistentes e à correção de valores em falta, através da remoção dos mesmos com ajuda do *Pentaho* e de *scripts* em *Python*

Durante a construção do sistema, deparámo-nos com alguns problemas, como já referido anteriormente, relacionados principalmente com dados mal formatados e vazios. Para resolver estes problemas, o grupo teve de aplicar uma normalização para tornar os dados todos uniformes. Um exemplo significativo dessa normalização ocorreu com os valores da temperatura, dos quais, à primeira vista, pareciam absurdos, variando de  $-20^{\circ}$  a  $110^{\circ}\text{C}$ . Ao analisar esses dados, o grupo procurou na descrição do *dataset* por alguma explicação e constatou que o proprietário do *dataset* havia esquecido de multiplicar todos os valores da temperatura por 0.1

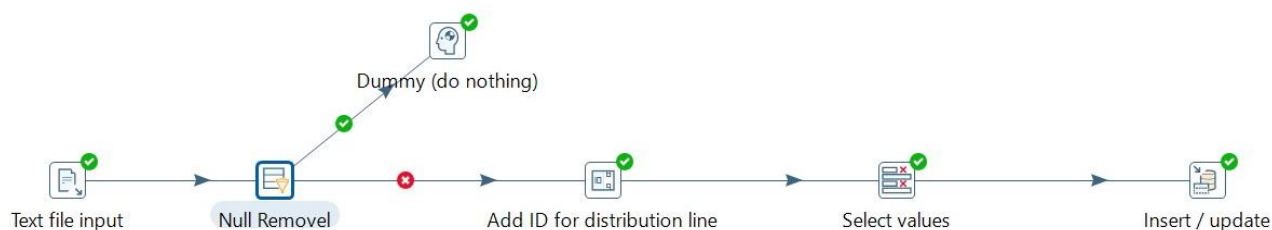
Destacamos que a limpeza dos dados meteorológicos exigiu uma abordagem mais criteriosa, devido à presença de informações datadas e irrelevantes para o nosso estudo e após a limpeza, realizámos a agregação de dados relevantes. Destacando mais uma vez os dados meteorológicos, realizou-se a agregação calculando as médias para várias variáveis associadas a dias específicos ao longo dos anos. Adicionalmente, aplicamos uma normalização aos dados numéricos, convertendo-os para decimais, exceto para o campo referente ao ano.

Por fim, recorreremos às funcionalidades da ferramenta *Pentaho* para o carregamento dos dados transformados para a nossa base de dados *PostgreSQL*. O *Pentaho* revelou-se uma escolha acertada, oferecendo-nos mecanismos eficazes para automatizarmos este processo de integração, garantido a qualidade e consistência dos dados armazenados.



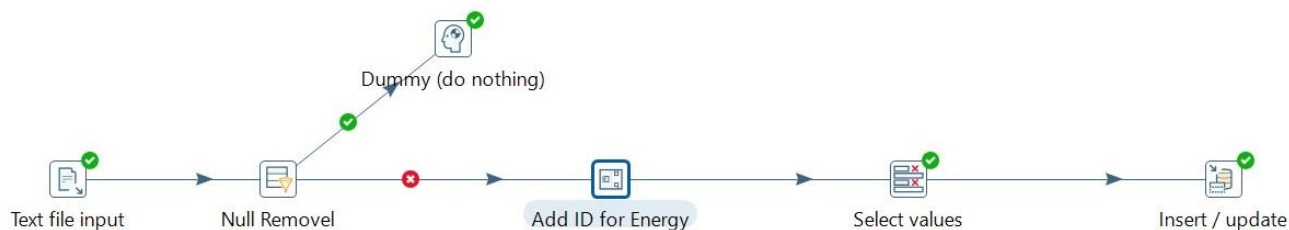
*Figura 4 - Transformação do dataset weather*

Para a transformação apresentada na figura 4, pelo facto de que o *dataset* das empresas serem muito extensos e por indicação do *Pentaho*, decidiu-se aplicar uma condição de que se existisse algum valor *null* em alguma linha do nosso *dataset*, essa linha seria retirada. Para finalizar, a transformação para a tabela do *weather* nos Países Baixos, foi preciso por um processo de seleção de algumas das *features* do *dataset* respetivas à tabela para que depois fossem inseridas na nossa *Data Warehouse*.



**Figura 5 - Transformação do dataset da energia consumida para a tabela de linha de distribuição.**

Para a transformação apresentada na figura 5, o mesmo teve de ser repetido, mas agora para a tabela de factos a partir do mesmo *dataset*.

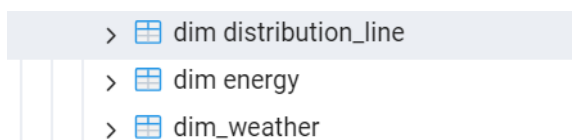


**Figura 6 - Transformação do dataset da energia consumida para a tabela de factos.**

## 8. OLAP

Neste capítulo, pretendemos demonstrar como organizámos e procedemos na secção do *OLAP*.

O nosso sistema de *OLAP* apresenta três dimensões específicas: o consumo de energia (que é o nosso facto principal), as condições meteorológicas em que ocorreu esse consumo e a linha de distribuição em que ocorreu esse mesmo consumo. E por esse mesmo motivo, e tendo em consideração o nosso modelo estrela, o *Pentaho* criou as três tabelas respetivas, conforme a figura mostra:



**Figura 7 – Tabelas criadas por o Pentaho relativas às dimensões**

```
1 SELECT dl.city, SUM(e.delivery_perc) AS total_delivery_percentage
2 FROM "dim distribution_line" dl
3 JOIN "dim energy" e ON dl.distribution_line_id = e.distribution_line_id
4 GROUP BY dl.city;
```

**Figura 8 – Query que retorna a soma da percentagem de energia que foi servida por cidade**

```
1 SELECT w.ug, SUM(e.annual_consume) AS total_annual_consume
2 FROM "dim energy" e
3 JOIN dim_weather w ON e.year = w.yyyymmdd
4 GROUP BY w.ug;
```

**Figura 9 – Query que retorna a soma anual do consumo e a média da humidade, por ano**

## 9. ESQUEMAS E MOCKUPS DO POWER BI

Nesta secção, apresentamos os vários *mockups* usados durante o projeto. Estes *mockups* foram e são importantes tanto para quem está a contruir uma solução de Inteligência no Negócio e para quem se vai apresentar estas soluções pois facilitam a construção das mesmas bem como facilitam a visualização e compreensão dos nossos *stakeholders*.

Para a elaboração dos *mockups*, utilizamos a ferramenta *PowerBI* e configurámos cada um deles de modo a apresentar uma quantidade relevante de informação aos nossos *stakeholders*, dando-lhes a possibilidade de tomar decisões estratégicas com base em dados concretos.

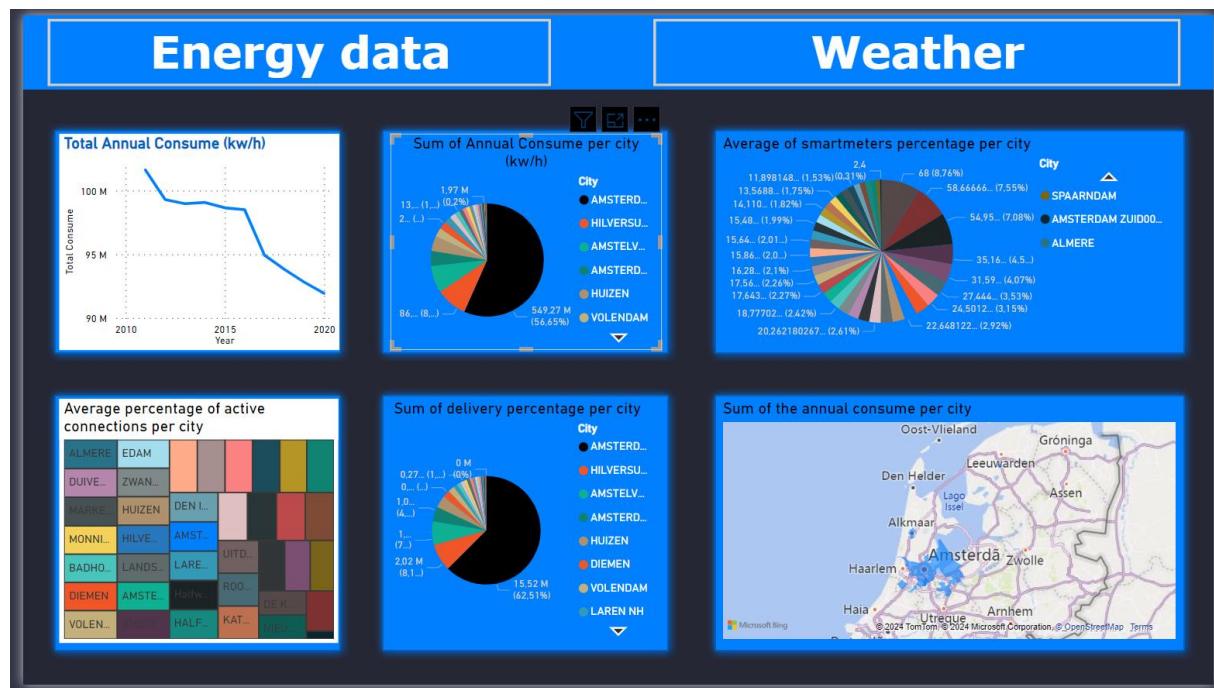


Figura 10 – Mockup “Energy Data”

Este *mockup* representa um conjunto gráficos alusivos a várias variáveis sobre:

1. Total anual consumido em (Kwh)
2. Soma do consumo anual por cidade (Kwh)
3. Média de contadores inteligentes por cidade
4. Percentagem média de conexões ativas por cidade
5. Soma da percentagem de entrega por cidade
6. Soma do consumo anual por cidade

Os dados apresentados nos *mockups* relativos às variáveis energéticas são fundamentais para os nossos *stakeholders*. Estes oferecem uma visão crucial sobre o consumo e distribuição da energia, permitindo que sejam feitas decisões estratégicas e a implementação de medidas eficazes.

O total anual consumido (Kwh) fornece uma visão geral do consumo de energia ao longo do ano, enquanto a Soma do Consumo Anual por cidade (Kwh) permite uma análise mais detalhada das diferenças de consumo entre regiões. As métricas relacionadas aos contadores inteligentes e conexões ativas destacam a eficiência da infraestrutura tecnológica das redes elétricas, enquanto a percentagem de entrega por cidade avalia a eficiência da distribuição de energia.

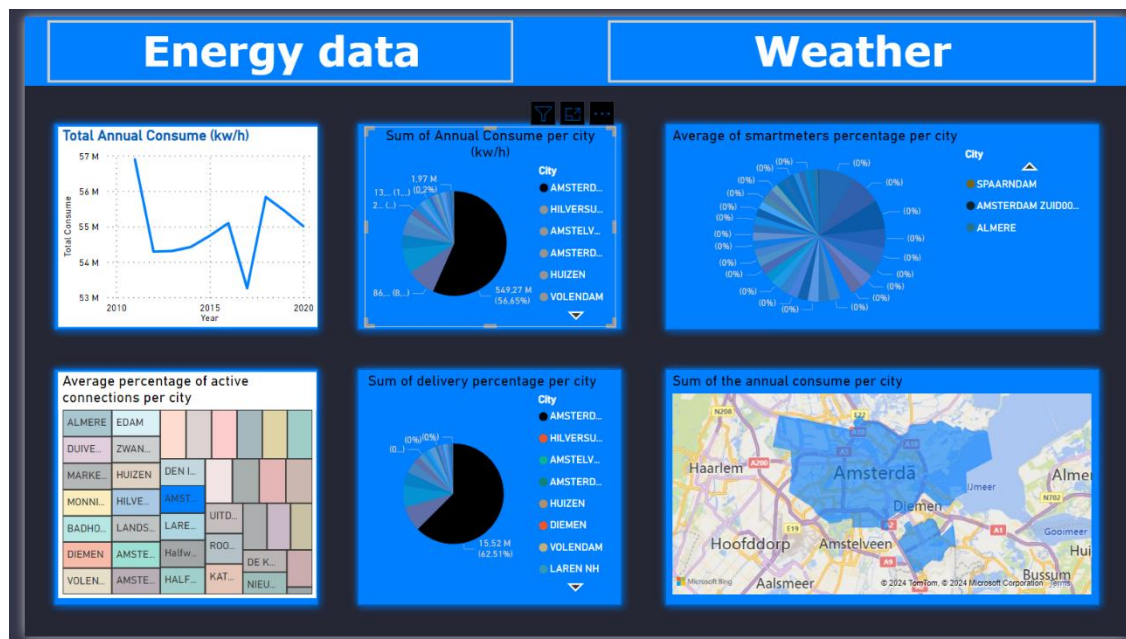


Figura 11 – Mockup “Energy Data”

Se o *stakeholder* pretender obter mais informações sobre uma cidade específica, ele pode fazê-lo clicando num dos gráficos que representam informações relativas a cidades. Essa interação permite uma exploração mais detalhada dos dados relativos à cidade selecionada, oferecendo uma experiência mais personalizada e informativa aos utilizadores.

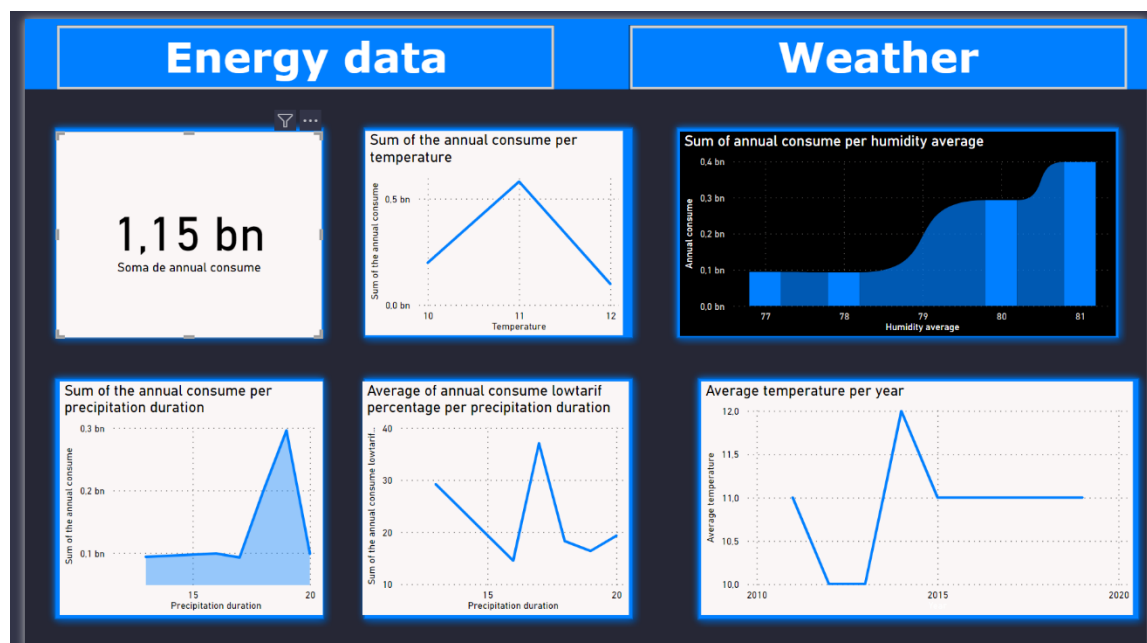


Figura 12 – Mockup “Weather Data”

Na segunda parte do sistema, os dados relacionados ao consumo de energia são combinados com informações meteorológicas dos Países Baixos. Nesta página, o **stakeholder** pode examinar a relação entre a meteorologia e o consumo de energia, como a média da temperatura num ano específico e como tal afetou o consumo da energia. Além disso, o utilizador pode realizar uma análise mais detalhada ao clicar num gráfico, uma vez que todos os restantes são dinâmicos o suficiente para se alterarem conforme a interação do utilizador. Esta funcionalidade permite uma exploração mais aprofundada dos padrões e tendências presentes nos dados, fornecendo uma compreensão mais abrangente dos fatores que influenciam o consumo de energia

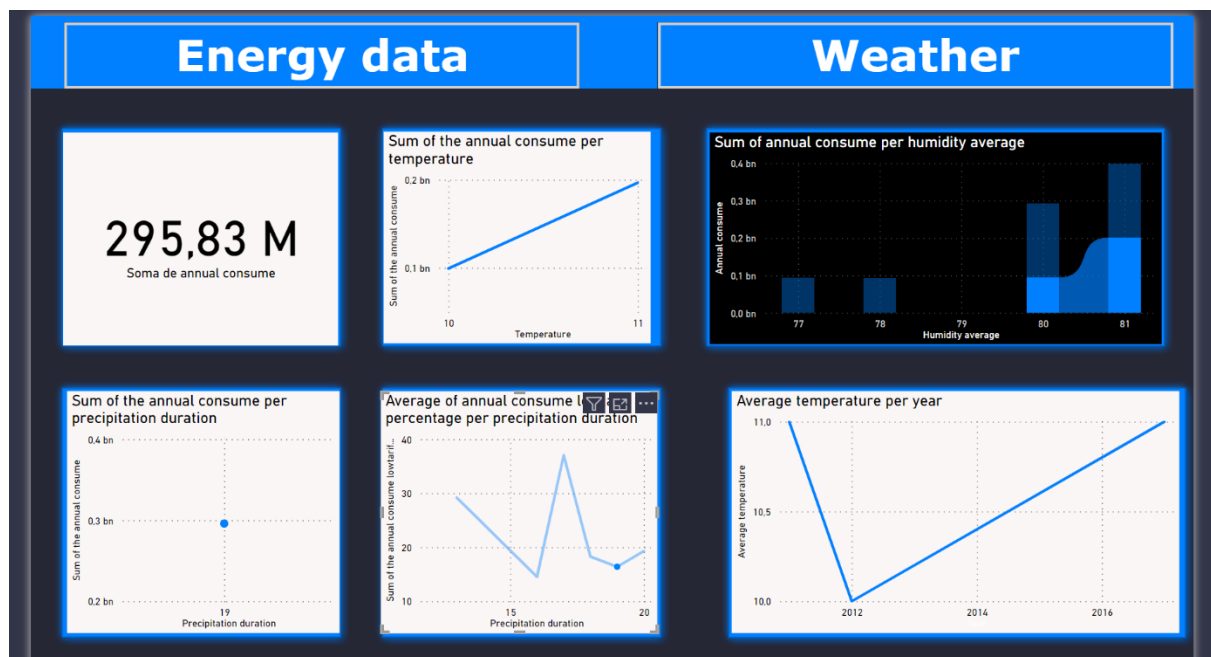


Figura 13 – Mockup “Weather Data”

## 10. CONCLUSÃO

Em suma, concluímos que o grupo cumpriu com sucesso em apresentar uma solução que esteja alinhada com os princípios fundamentais de *BI (Business Intelligence)*, como delineado ao longo do presente relatório.

A análise dos dados recolhidos emerge como um elemento crucial para toda a conceptualização do sistema e para além disso, serve como teste do mesmo, visto que o nosso sistema é responsivo para todo o tipo de empresas que distribuam energia.

Quanto às tecnologias utilizadas, discutimos as escolhas fundamentais relacionadas à arquitetura de dados, bem como todos os processos essenciais e relacionados com *ETL* e as soluções *OLAP*, destacando especialmente a sua relevância na análise dinâmica e interativa dos dados.

A apresentação dos *mockups* proporcionou uma visão concreta da possível solução desenvolvida, permitindo uma compreensão clara das funcionalidades chave e da interação dos *stakeholders* com o sistema.

Com esta combinação de elementos, formou-se uma base sólida para a aplicação eficiente de *BI*, dando a capacidade às organizações a extrair insights valiosos e a tornar decisões fundamentadas com base em evidências sólidas.

Por fim, a nossa solução oferece uma abordagem holística e orientada a bons resultados, com o potencial de impulsionar o sucesso e a competitividade das empresas no mercado atual.

## 11. BIOGRAFIA

<https://www.kaggle.com/datasets/lucabasa/dutch-energy/data>

<https://www.kaggle.com/code/kerneler/starter-knmi-1901-2020-historical-d155b8c0-7/notebook>

<https://www.microsoft.com/pt-pt/power-platform/products/power-bi> - PowerBI

[https://www.google.com/urlsa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi\\_36\\_8wfiEAxUU1wIHHVhXBksQFnoECACQAQ&url=https%3A%2F%2Fwww.hitachivantara.com%2Fen-us%2Fproducts%2Fpentaho-plus-platform%2Fdata-integration-analytics.html&usg=AOvVaw1d5TmmH9B7ys\\_ZzbQm9BJc&opi=89978449](https://www.google.com/urlsa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi_36_8wfiEAxUU1wIHHVhXBksQFnoECACQAQ&url=https%3A%2F%2Fwww.hitachivantara.com%2Fen-us%2Fproducts%2Fpentaho-plus-platform%2Fdata-integration-analytics.html&usg=AOvVaw1d5TmmH9B7ys_ZzbQm9BJc&opi=89978449) - Pentaho

<https://www.postgresql.org> - PostgreSQL

<https://www.ibm.com/topics/etl> - ETLs