

Solução de Business Intelligence aplicado a incêndios florestais

Eliana Oliveira
Mestrado em Ciência dos dados
Instituto Politécnico de Leiria - ESTG
Leiria, Portugal
2240276@my.ipleiria.com

Jéssica Grácio
Mestrado em Ciência dos dados
Instituto Politécnico de Leiria - ESTG
Leiria, Portugal
2240549@my.ipleiria.pt

Maria João Fialho
Mestrado em Ciência dos dados
Instituto Politécnico de Leiria - ESTG
Leiria, Portugal
2240286@my.ipleiria.pt

Resumo— O presente relatório descreve o desenvolvimento de um projeto de *Business Intelligence* centrado na análise dos incêndios florestais em Portugal Continental entre os anos de 2011 e 2020. Através da utilização de dados oficiais disponibilizados pelo Instituto da Conservação da Natureza e das Florestas (ICNF), foi implementado um processo completo de BI, incluindo a modelação dimensional, integração de dados, análise e visualização de indicadores relevantes. O principal objetivo é identificar padrões temporais e espaciais nos incêndios florestais, analisar a sua gravidade e frequência, e apoiar decisões estratégicas no domínio da prevenção e combate. A solução desenvolvida inclui *dashboards* interativos e uma narrativa baseada em dados, direcionada a decisores públicos e técnicos da área florestal.

Palavras-chave —*Business Intelligence, Incêndios Florestais, ICNF, Portugal, Análise de Dados, Data Warehouse, Dashboards, Modelação Dimensional.*

I. INTRODUÇÃO

Nos últimos anos, os incêndios florestais têm-se tornado uma preocupação crescente em Portugal, não apenas pelo impacto ambiental significativo, mas também pelas consequências económicas e sociais que acarretam. A análise dos dados históricos de incêndios florestais é essencial para compreender os padrões e tendências subjacentes, identificar fatores críticos de risco, otimizar recursos de combate e apoiar políticas públicas mais eficazes de prevenção e mitigação.

Este projeto, desenvolvido no âmbito da unidade curricular de *Business Intelligence* do Mestrado em Ciência de Dados (2024/2025), tem como objetivo aplicar metodologias e ferramentas de Business Intelligence ao estudo dos incêndios florestais registados em Portugal continental no período compreendido entre 2011 e 2020. A base de dados utilizada foi obtida a partir do Sistema de Gestão de Informação de Incêndios Florestais (SGIF), disponibilizada pelo Instituto da Conservação da Natureza e das Florestas [1] (ICNF).

Ao longo deste projeto, serão percorridas todas as fases típicas de um projeto de *Business Intelligence*, desde a identificação das fontes de dados e levantamento de requisitos, até à conceção e implementação de um *data warehouse*, análise multidimensional dos dados, construção de *dashboards* interativos e desenvolvimento de uma narrativa orientada a dados. O principal objetivo é extrair conhecimento que ajude na compreensão do fenómeno dos incêndios e apoiar a tomada de decisão das entidades relevantes.

II. DESCRIÇÃO DOS OBJETIVOS

O objetivo principal deste projeto é aplicar metodologias de *Business Intelligence* à análise de incêndios florestais em Portugal continental, entre 2011 e 2020, com base nos dados

fornecidos pelo Sistema de Gestão de Informação de Incêndios Florestais (SGIF). Para tal, propõe-se um conjunto de etapas que abrangem desde a modelação dos dados até à visualização interativa da informação.

- Conceber e implementar um *data warehouse* orientado para o estudo dos incêndios com base nas dimensões de tempo, localização e causa.
- Responder a um conjunto de questões de análise predefinidas, com foco na área ardida, número de incêndios e padrões temporais e espaciais.
- Construir *dashboards* interativos no *Power BI* que permitam explorar a informação de forma dinâmica, clara e intuitiva.
- Identificar padrões como as épocas do ano mais críticas, locais com maior número de incêndios, principais causas e comportamento dos incêndios ao longo do tempo.

Com este projeto, pretende-se não só aprofundar o conhecimento sobre os incêndios florestais em Portugal, mas também fornecer ferramentas analíticas que possam apoiar a tomada de decisões mais eficazes na prevenção e combate destes fenómenos.

III. DESCRIÇÃO DOS PROMOTORES

A entidade promotora do projeto é o Instituto da Conservação da Natureza e das Florestas (ICNF), organismo público responsável pela gestão da informação sobre incêndios florestais em Portugal. O ICNF disponibiliza dados oficiais através do Sistema de Gestão de Informação de Incêndios Florestais (SGIF), servindo de base à análise desenvolvida neste projeto.

A equipa responsável pelo desenvolvimento da solução é constituída por Eliana Oliveira, Jéssica Grácio e Maria Fialho e no papel de analistas e consultoras externas, responsáveis pela conceção, modelação e implementação de um sistema de *Business Intelligence* que permita responder a um conjunto de questões relacionadas com os incêndios florestais.

IV. ENTREVISTAS

Para garantir que o sistema de *Business Intelligence* desenvolvido responde de forma eficaz às necessidades reais de análise no contexto dos incêndios florestais, foram realizadas entrevistas simuladas com representantes de entidades relevantes nesta área. Estas entrevistas tiveram como objetivo identificar expectativas, problemas operacionais e necessidades estratégicas das organizações que atuam na prevenção, combate e gestão dos incêndios.

Com base nas declarações recolhidas durante estas entrevistas, foram construídas *user stories* que refletem as necessidades, expectativas e desafios sentidos pelos

utilizadores. Estas narrativas ajudaram a orientar a definição dos requisitos funcionais, com foco na criação de valor para os utilizadores finais.

Abaixo apresentam-se os testemunhos das entidades entrevistadas e as respetivas *user stories* extraídas de cada uma.

A. Entrevista com o Diretor do Departamento de Prevenção de Incêndios do ICNF:

“Para nós, no ICNF, é essencial perceber o impacto dos incêndios em termos de área ardida e como este fenómeno tem evoluído ao longo dos anos. Precisamos de dados fiáveis que nos ajudem a entender os meses mais críticos e as principais causas das ocorrências. Com base nisso, podemos reforçar medidas de prevenção e sensibilização.”

Tabela 1 - Entrevista 1

User Stories (Diretor do Departamento de Prevenção de Incêndios do ICNF)	
Título: US1 – Visualizar a área total ardida por ano e por mês.	
Descrição: Como diretor do ICNF, pretendo visualizar a área total ardida em cada ano e em cada mês para avaliar a gravidade dos incêndios ao longo do tempo.	
Título: US2 – Visualizar o número de incêndios por causa	
Descrição: Como diretor do ICNF, pretendo identificar o número de incêndios por causa para conhecer os principais fatores de risco.	
Título: US3 – Visualizar os meses em que ocorrem mais incêndios	
Descrição: Como diretor do ICNF, pretendo consultar os meses com mais incêndios para reforçar ações de prevenção nesses períodos.	

B. Entrevista ao Comandante Distrital da Proteção Civil

“A nossa preocupação no terreno passa por saber quando e onde os incêndios acontecem com mais frequência. Precisamos de ter uma visão clara dos concelhos e distritos mais afetados, assim como saber em que alturas do dia ou da semana e quantas ocorrências ocorrem ao longo do ano. Isso permite-nos alocar meios de forma mais eficaz, reforçar equipas e otimizar a resposta operacional.”

Tabela 2 - Entrevista 2

User Stories (Comandante Distrital da Proteção Civil)	
Título: US4 – Visualizar as zonas do país com mais incêndios	
Descrição: Como comandante distrital da Proteção Civil, pretendo identificar os distritos e concelhos com mais incêndios para reforçar a prontidão local.	
Título: US5 – Identificar os períodos do dia com mais incêndios	
Descrição: Como comandante distrital da Proteção Civil, pretendo saber em que períodos do dia há mais ignições para reforçar as equipas nesses horários.	
Título: US6 – Visualizar o número de incêndios por ano	
Descrição: Como responsável pelo planeamento estratégico, pretendo visualizar o número total de incêndios por ano para analisar a evolução da frequência de ocorrências ao longo do tempo e identificar tendências de crescimento ou redução	

C. Entrevista com o Coordenador Nacional da Força Especial de Bombeiros (FEB)

“Do ponto de vista do combate direto, interessa-nos perceber a duração média dos incêndios e identificar os casos que ultrapassam as 24 horas. Estas ocorrências exigem grande esforço logístico e coordenação nacional. Com essa informação conseguimos planejar melhor a mobilização de meios.”

Tabela 3 - Entrevista 3

User Stories (Coordenador Nacional da Força Especial de Bombeiros (FEB))	
Título: US7 – Visualizar a duração média dos incêndios	
Descrição: Como coordenador nacional da FEB, pretendo conhecer a duração média dos incêndios para planejar os recursos de combate.	
Título: US8 – Visualizar os incêndios com duração superior a 24 horas	
Descrição: Como coordenador nacional da FEB, pretendo identificar os períodos em que ocorrem mais incêndios longos para melhorar a capacidade de resposta prolongada.	

V. REQUISITOS

A definição dos requisitos funcionais é uma etapa essencial num projeto de *Business Intelligence*, uma vez que orienta o processo da *data warehouse* e garante que o sistema final responde às necessidades reais de análise e tomada de decisão. Os requisitos funcionais estabelecidos, representam as funcionalidades esperadas do sistema, nomeadamente ao nível da visualização e interpretação dos dados no *Power BI* [5].

A tabela seguinte apresenta os requisitos identificados, bem como a sua descrição e a prioridade atribuída a cada um.

Tabela 4 - Definição de Requisitos

	Designação	Descrição	Prioridade
RF 01	Visualizar a área total ardida por ano e por mês	O sistema deve permitir visualizar a área total ardida em cada ano e cada mês.	Elevada
RF 02	Visualizar o número de incêndios por causa	O sistema deve permitir identificar o número de incêndios associados a cada tipo de causa.	Elevada
RF 03	Visualizar o número de incêndios por ano	O sistema deve permitir analisar a evolução do número de incêndios ao longo dos anos.	Elevada
RF 04	Visualizar os meses em que ocorrem mais incêndios	O sistema deve permitir consultar a o número de incêndios em cada mês.	Elevada
RF 05	Visualizar as zonas do país com mais incêndios	O sistema deve permitir analisar o número de incêndios por zona.	Elevada
RF 06	Visualizar os períodos do dia com mais tendência ao inícios de incêndios	O sistema deve permitir identificar os períodos do dia com maior tendência a que se iniciem incêndios.	Média
RF 07	Visualizar a duração média dos incêndios	O sistema deve permitir calcular a duração média dos incêndios, sempre que essa informação existir.	Média
RF 08	Visualizar os incêndios com duração > 24h	O sistema deve permitir identificar em que meses ocorrem mais incêndios com duração superior a 24h.	Média

VI. QUESTÕES A RESPONDER

Foram definidas um conjunto de questões relevantes que visam guiar a construção dos *dashboards* no *Power BI* e permitir uma análise clara e útil do fenómeno dos incêndios florestais em Portugal entre 2011 e 2020.

Estas questões foram formuladas com base nos objetivos do projeto, nas entrevistas realizadas com os *stakeholders* e nas necessidades analíticas identificadas durante o levantamento de requisitos. O seu principal propósito é extrair conhecimento relevante a partir dos dados e apoiar a tomada

de decisão por parte das entidades envolvidas na prevenção, combate e gestão de incêndios.

Abaixo apresenta-se a lista das questões analíticas a responder:

1. Qual foi a área total ardida em cada ano?
2. Qual é o número de incêndios por causa?
3. Quantos incêndios ocorreram por ano?
4. Qual é a área ardida por distrito?
5. Qual é a duração média dos incêndios?
6. Quais são os distritos com mais incêndios?
7. Em que altura do dia ocorrem mais incêndios?
8. Quais os meses do ano em que ocorrem mais incêndios?
9. Quantas ocorrências apresentaram uma duração superior a 24 horas?
10. Quais são as áreas mais afetadas: agrícolas ou povoadas?
11. Qual foi a principal causa de incêndios no ano com maior número de ocorrências?
12. Quais são as cidades ou distritos com maior área ardida?

As respostas a estas questões serão apresentadas através de visualizações e *dashboards* desenvolvidos no *Power BI*. Todas as análises podem ser consultadas na secção XI - Análise e narrativa dos dados.

VII. FONTES DE DADOS ORIGINAIS

Para este projeto, foi utilizada como fonte principal de dados os ficheiros disponibilizados pelo Instituto da Conservação da Natureza e das Florestas [1] (ICNF), especificamente os dados de incêndios florestais entre os anos de 2011 a 2020.

O ficheiro original encontrava-se em formato .csv e continha um total de 41 variáveis, incluindo informações espaciais, meteorológicas, causas, localização, e tempo de ocorrência dos incêndios.

Contudo, para construir um modelo dimensional simplificado e adequado à análise, foi realizado um processo de limpeza e transformação dos dados, utilizando o Microsoft Excel como ferramenta principal, onde foram implementadas as seguintes decisões:

1. Foram removidas colunas relacionadas com coordenadas geográficas (como latitude, longitude, X/Y militar, ETRS89), uma vez que não eram relevantes para as análises pretendidas.
2. Foram mantidas apenas os atributos essenciais para construir as dimensões de causa, tempo e local, bem como as medidas necessárias para a tabela de factos.
3. Foram criados ficheiros auxiliares separados para as tabelas Dim_Causa e Dim_Local, com as colunas relevantes agregadas a partir do ficheiro principal.
4. Foram removidos registos de incêndios que apresentavam exatamente a mesma data e hora (até ao segundo) de início e fim, uma vez que esta coincidência perfeita foi considerada um provável erro humano. Além disso, como a

combinação da data e hora de início e fim foram utilizadas para construir a chave primária no processo de integração dos dados, foi necessário eliminar estes duplicados para garantir a consistência e unicidade do identificador.

Foram também construídos ficheiros auxiliares – tabelas de dimensão - Dim_Data, Dim_Hora, Dim Local e Dim Causa.

A Dim Data contém os anos entre 2000 e 20240, caso haja necessidade de inserir incêndios passados ou futuros e foi criada para permitir análise temporal ao nível do dia, mês, trimestre, semestre e ano.

A Dim_Hora foi derivada da hora do alerta, permitindo análises por hora, minuto e período do dia (madrugada, manhã, tarde e noite). Estas dimensões são fundamentais para suportar funcionalidades de análise temporal e identificar padrões de ocorrência dos incêndios ao longo do tempo.

Por sua vez, a Dim Local e Dim Causa foram criadas com base na base de dados original. As colunas referentes aos locais e causas foram removidos do ficheiro original, mantendo apenas o código correspondente e foram copiados para os ficheiros auxiliares. No caso da Dim_Local, identificaram-se duplicados associados ao mesmo código de local, onde o distrito permanecia igual, mas os campos de concelho e freguesia apresentavam variações ou erros de escrita. Por exemplo, o código “10101” surgia com o distrito “Aveiro”, mas aparecia associado tanto ao concelho “Agueda” como a “Águeda”. Esta inconsistência foi resolvida durante o processo de limpeza, de forma a assegurar que cada código de local fosse único e corretamente normalizado.

A. Modelo de Dados Originais

A Ilustração 1 apresenta o modelo relacional que resulta da estrutura dos dados originais. Este modelo organiza os dados em torno da tabela central Incendios, que contém as principais informações sobre cada ocorrência. A ligação com as tabelas Causa e Local permite obter detalhes sobre a origem e a localização dos incêndios, respetivamente. As tabelas Data e Hora também são representadas para demonstrar a futura separação dos dados temporais no modelo dimensional.

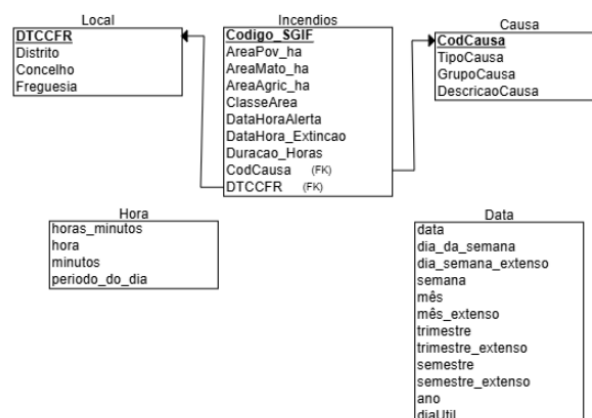


Ilustração 1 - Modelo de Dados Originais

B. Identificação e estruturação de factos e dimensões

Para suportar a construção do modelo dimensional, foi elaborada uma matriz de associação entre os factos e as dimensões, que permitiu visualizar a estrutura das dependências dimensionais de cada medida de negócio.

Esta matriz evidencia as tabelas de dimensão que contextualizam cada um dos factos presentes na tabela central Incendio_Fact. Esta matriz foi construída com base num *template* que tinha sido disponibilizado nas aulas, o que assegurou o alinhamento com os conceitos lecionados e facilitou a identificação das dependências dimensionais.

Abaixo, na Ilustração 2, apresenta-se a Matriz de Utilização de Dimensões, onde o valor 1 indica a ligação entre um facto e a respetiva dimensão:

	Factos	Dimensões					Contagem da Utilização das Dimensões
		Data	Hora	Local	Causa		
Incêndios	AreaPov_ha	1	1	1	1	1	4
	AreaMato_ha	1	1	1	1	1	4
	AreaAgric_ha	1	1	1	1	1	4
	AreaTotal_ha	1	1	1	1	1	4
	ClasseArea	1	1	1	1	1	4
	Duracao_Horas	1	1	1	1	1	4
	IncSup24horas	1	1	1	1	1	4
	Contagem: Utilização dos Factos	7	7	7	7	7	

Ilustração 2 - Matriz de Relacionamentos entre a tabela de Facto e Dimensões

A estrutura detalhada de cada tabela (factos e dimensões), com os seus atributos, regras de negócio e ligações será explorada no capítulo VIII. Mapeamento entre dados originais e destino.

C. Diagrama do modelo de dados dimensional

A estrutura do modelo dimensional foi desenhada segundo o modelo estrela (*Star Schema*), no qual a tabela de factos “Incendio_fact_table” está no centro, ligada a várias tabelas de dimensão que fornecem contexto às ocorrências de incêndio florestal.

As dimensões foram estruturadas para permitir a análise sob diferentes perspetivas, nomeadamente: tempo (data e hora), causa, e localização.

De forma a permitir análises temporais mais granulares e comparações entre momentos distintos do processo do incêndio (alerta e extinção), optou-se por representar a mesma dimensão (Data e Hora) com duas instâncias distintas:

- Data_Inicio e Hora_Inicio: relacionadas com o momento de alerta do incêndio.
- Data_Fim e Hora_Fim: relacionadas com o momento de extinção do incêndio.

Esta abordagem permite, por exemplo, avaliar a duração dos incêndios com base em data e hora, ou comparar padrões de início e fim ao longo do tempo, sem comprometer a integridade do modelo.

A tabela de factos armazena também atributos como:

- Áreas ardidas por tipo (povoamento, mato, agrícola e total),
- Duração do incêndio,
- Classificação da área ardida (ClasseArea),
- Indicador binário de incêndios com duração superior a 24 horas.

Entre os atributos da tabela de factos, destacam-se duas colunas calculadas. A *AreaTotal_ha* é obtida pela soma das

áreas ardidas em povoamentos florestais, matos e zonas agrícolas, o que permite uma visão agregada da área total afetada. Já a variável *IncSup24Horas* é um indicador binário que assume o valor 1 sempre que a duração do incêndio ultrapassa as 24 horas, sendo calculada com base no atributo *Duracao_Horas*.

Cada uma das dimensões foi construída com base nas colunas do ficheiro original e complementada com colunas derivadas e regras de negócio (por exemplo, identificação de período do dia ou se o dia é útil).

A estrutura final do modelo dimensional, com a tabela de factos e as respetivas dimensões associadas, está representada na Ilustração 3.

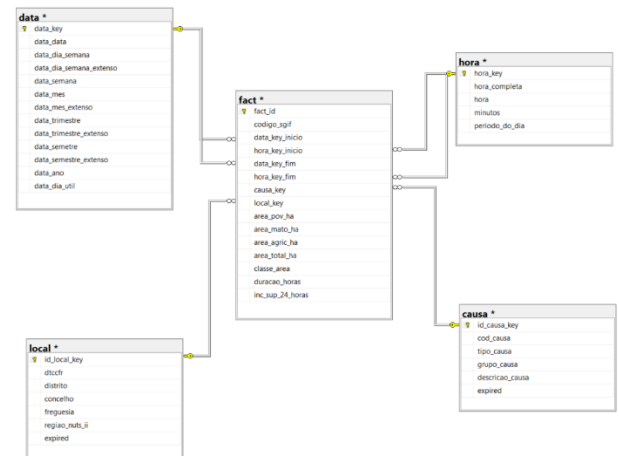


Ilustração 3 - Modelo Dimensional

VIII. MAPEAMENTO ENTRE DADOS ORIGINAIS E DESTINO

Para garantir a coerência e rastreabilidade dos dados utilizados na construção do modelo dimensional, foi realizado um mapeamento entre os atributos dos ficheiros originais (disponibilizados pelo ICNF) e as tabelas de destino, que compõem o *data warehouse*.

Esse mapeamento permitiu:

1. Identificar quais colunas foram mantidas, transformadas ou descartadas.
2. Associar atributos aos respetivos papéis no modelo (factos ou dimensões).
3. Documentar regras de derivação e domínios aplicados.

De forma a garantir a rastreabilidade completa dos dados, foram elaboradas tabelas de mapeamento que detalham, para cada atributo, a sua origem no ficheiro original, o respetivo papel no modelo dimensional (facto ou dimensão), se foi transformado ou mantido, e as regras de derivação aplicadas.

Devido à dimensão desta informação, esses mapeamentos encontram-se organizados nos Anexos A a E, um por cada tabela do modelo (Incêndio, Data, Hora, Local e Causa), o que permitiu uma consulta mais clara e detalhada.

IX. IMPLEMENTAÇÃO DO DATA WAREHOUSE

A implementação do *Data Warehouse* foi realizada no Visual Studio 2022 [4], através da plataforma SQL Server Integration Services (SSIS) [3]. O processo seguiu as boas práticas de projetos ETL, e envolveu a extração,

transformação e carregamento de dados provenientes de fontes externas para o modelo dimensional previamente definido.

O *Control Flow* foi desenhado para garantir a execução sequencial das tarefas. A execução iniciou-se com o carregamento das tabelas de dimensão: Data, Hora, Causa e Local. Cada uma destas tarefas foi desenvolvida com componentes do tipo *Data Flow*, nos quais foram aplicadas as transformações necessárias, derivaram-se atributos, implementaram-se regras de negócio e assegurou-se a qualidade dos dados antes do seu armazenamento no *data warehouse*.

Após o carregamento das dimensões, executou-se uma tarefa SQL (*Execute SQL Task*), que preparou o ambiente para o carregamento da tabela de factos, de forma a assegurar que apenas novos incêndios são inseridos na tabela de factos.

Por fim, a tabela de factos foi preenchida através de um *Data Flow* dedicado. Este integrou os dados dos incêndios com as respetivas chaves (*surrogate keys*) das dimensões, o que possibilitou a análise dos eventos com base nos seus atributos temporais, geográficos e causais. A estrutura do pacote SSIS desenvolvido encontra-se representada na figura abaixo:

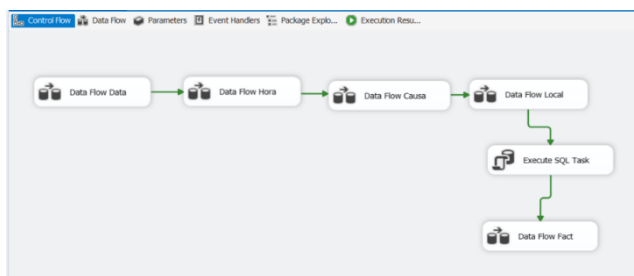


Ilustração 4 - Processo ETL implementado no SSIS

Este processo permitiu a criação de um *Data Warehouse* funcional, com todas as tabelas dimensionais e a tabela de factos devidamente integradas e preparadas para análises posteriores.

A seguir, descreve-se em detalhe o que foi realizado em cada um dos *Data Flows* apresentados, evidenciando as etapas de extração, transformação e carregamento específicas de cada dimensão e da tabela de factos.

A. Data Flow - Dimensões Temporais (Data e Hora)

O carregamento das dimensões Data e Hora foi efetuado a partir de ficheiros em formato Excel, contendo respetivamente os dados temporais relativos a datas e a horas/minutos.

O fluxo de dados para ambas as dimensões seguiu uma lógica comum, composta pelas seguintes etapas:

1. *Excel Source*: componente responsável por ler os ficheiros Excel com os atributos temporais relevantes.
2. *Data Conversion*: transformação aplicada para adaptar os tipos de dados aos formatos exigidos pelo SQL Server.
3. *Slowly Changing Dimension* (SCD): componente utilizado para identificar alterações em atributos definidos como fixos, assegurando que valores anteriormente carregados não são substituídos. Neste

caso, a lógica aplicada impede a substituição dos dados, mantendo os valores originais inalterados.

4. *Insert Destination*: caminho usado para inserir novos registos nas tabelas *dim_data* ou *dim_hora*, conforme a dimensão em questão.
5. *OLE DB Command*: módulo utilizado para atualizar registos existentes quando se detetaram alterações em atributos não identificadores.

Este processo assegurou a correta integração e atualização das dimensões temporais, garantiu a consistência da informação e evitou duplicações.

As representações visuais dos fluxos encontram-se nos Anexos F (*Data Flow Dimensão Data*) e G (*Data Flow Dimensão Hora*), onde é possível consultar uma visualização mais detalhada de cada etapa.

B. Data Flow Causa

A dimensão Causa foi construída com base num ficheiro Excel que continha informação relativa às causas dos incêndios. Esta dimensão exigiu a gestão de alterações em atributos históricos, pelo que se adotou um modelo *Slowly Changing Dimension* (SCD) Tipo 2, que preservou o histórico sempre que os dados sofreram modificações relevantes.

O fluxo de dados desta dimensão seguiu as seguintes etapas:

1. *Excel Source*: módulo que acedeu aos dados do ficheiro Excel com os atributos da dimensão Causa.
2. *Data Conversion*: transformação que assegurou a compatibilidade dos tipos de dados com os formatos exigidos pela base de dados de destino.
3. *Slowly Changing Dimension*: componente que tratou esta dimensão como SCD Tipo 2. Os campos GrupoCausa e DescricaoCausa foram classificados como atributos históricos, por estarem sujeitos a alterações manuais por erro humano. Sempre que o sistema deteta alterações nestes campos, cria registos para preservar o histórico. A eficácia desta configuração foi validada com um teste de atualização na freguesia, demonstrado no Anexo J – Carregamento Incremental na Dimensão Causa. (inclui-se no anexo uma imagem do passo realizado).
4. *Derived Column*: etapa que criou uma coluna auxiliar chamada *expired*, usada para monitorizar se o registo atual foi substituído por um mais recente.
5. *OLE DB Command*: módulo que atualizou os registos existentes após identificar alterações em atributos não identificadores.
6. *Union All*: operação que agregou os fluxos de inserções e atualizações num conjunto único de dados.
7. *Derived Column 1*: módulo que tratou colunas técnicas adicionais antes da carga final.
8. *Insert Destination*: componente responsável por inserir os dados finais na tabela *dim_causa*.

Este fluxo preservou o histórico da dimensão Causa, assegurou a atualização da informação e manteve a integridade dos dados carregados.

A representação visual deste fluxo encontra-se disponível no Anexo H – Data Flow Dimensão Causa, onde se apresenta a configuração detalhada desta componente do ETL.

C. Data Flow Local

A dimensão Local foi construída a partir de um ficheiro Excel com os dados geográficos dos incêndios. Esta dimensão também exigiu a preservação do histórico de alterações e seguiu um processo semelhante ao da dimensão Causa, o que também acabou por adotar o modelo *Slowly Changing Dimension* (SCD) Tipo 2.

O fluxo de dados incluiu os seguintes passos:

1. *Excel Source*: módulo que acedeu ao ficheiro Excel com os atributos geográficos.
2. *Derived Column* – Distrito: transformação aplicada para uniformizar os nomes dos distritos. Foi feita a correção do valor “Viana Do Castelo” foi convertido em “Viana do Castelo” para garantir o mesmo nome do distrito.
3. *Derived Column* – Região NUTS II: módulo que atribuiu a cada distrito uma região correspondente (Norte, Centro, Lisboa, Alentejo ou Algarve). A derivação baseou-se numa expressão condicional que classificou cada distrito com a sua região NUTS II [7], ou atribuiu “Desconhecido” quando o valor não correspondia a nenhum caso esperado.
4. *Data Conversion*: transformação que assegurou a compatibilidade dos tipos de dados com os formatos exigidos pela base de dados de destino.
5. *Slowly Changing Dimension*: componente que tratou a dimensão como SCD Tipo 2. Esta configuração permitiu preservar o histórico de alterações no atributo relevante, a freguesia, sobretudo após a reorganização administrativa das freguesias ocorrida em 2013, que alterou a estrutura territorial de várias regiões em Portugal [2]. A eficácia desta configuração foi validada com um teste de atualização na freguesia, demonstrado no Anexo K – Carregamento Incremental na Dimensão Causa.
6. *Derived Column* – *Expired*: etapa que introduziu uma nova coluna para marcar os registos expirados.
7. *OLE DB Command*: módulo que atualizou registos existentes sempre que o sistema detetou alterações em atributos não identificadores.
8. *Union All*: operação que reuniu os novos registos e os modificados num conjunto único de dados.
9. *Derived Column* 1: transformação que ajustou colunas técnicas antes da carga final.
10. *Insert Destination*: componente que inseriu os dados transformados na tabela *dim_local*.

Este processo assegurou a qualidade e a rastreabilidade da dimensão Local, permitindo a atualização controlada da informação geográfica e a manutenção de um histórico coerente.

A representação visual deste fluxo encontra-se disponível no Anexo H – Data Flow Dimensão Local, que apresenta o mapeamento completo das transformações efetuadas.

D. Execute SQL Task

Antes da execução do fluxo de dados responsável pelo carregamento da tabela de factos, foi utilizada uma tarefa do tipo *Execute SQL Task* para recuperar o último registo carregado. Esta tarefa executou a seguinte instrução SQL:

```
SELECT ISNULL(MAX(fact_id), 0) AS  
LAST_INCENDIO FROM dbo.fact
```

Este comando permitiu obter o maior valor atual da chave primária *fact_id* na tabela *fact*, atribuindo o valor 0 nos casos em que a tabela se encontrasse vazia. O resultado foi armazenado numa variável que serviu como ponto de referência para identificar e carregar apenas os novos registos de incêndios, evitando duplicações e assegurando a integridade dos dados no processo incremental.

Para validar a funcionalidade da tarefa *Execute SQL Task*, realizou-se um teste prático com o objetivo de confirmar se o carregamento incremental da tabela de factos estava a funcionar corretamente. Antes da execução do fluxo de dados responsável por inserir novos registos, foi executada a instrução `SELECT COUNT(*) FROM [dw_incendios].[dbo].[fact]`, que devolveu um total de 176271 registos. Após a execução do processo de ETL com a inserção de um novo incêndio, a mesma consulta retornou o valor 176272, comprovando assim que a lógica implementada identificou corretamente o último registo carregado e adicionou apenas os dados novos. A evidência deste teste encontra-se representada no Anexo L – Carregamento Incremental na tabela de Factos.

E. Data Flow Fact

O processo de integração dos dados factuais relativos aos incêndios florestais incluiu a junção das informações provenientes das dimensões *Causa* e *Local* com os registos de factos obtidos a partir de um ficheiro Excel. O objetivo consistiu em construir os identificadores e indicadores necessários para o carregamento da tabela de factos do *Data Warehouse*.

A sequência de operações envolveu os seguintes passos:

1. *Excel Source*: módulo que acedeu ao ficheiro Excel com os dados dos incêndios.
2. *Derived Columns*: transformação responsável pela criação de várias colunas auxiliares. As expressões definidas permitiram:
 - Extrair a data e a hora de início e fim do incêndio, a partir dos campos *DataHoraAlerta* e *DataHoraExtincao*, para gerar os identificadores *data_key_inicio*, *data_key_fim*, *hora_key_inicio* e *hora_key_fim*.
 - Calcular a área total ardida (*area_total_ha*), com base nas colunas de área pública e privada.
 - Identificar incêndios com duração igual ou superior a 24 horas (*inc_sup_24_horas*).
 - Corrigir eventuais valores nulos no campo *CodCausa*, substituindo-os por zero.

3. **Derived Column – DataHora_ID:** criação de um identificador numérico único (*datahora_id*) que concatena a data e a hora de início e fim do incêndio. Esta coluna foi criada como alternativa à utilização do campo *Codigo_SGIF*, que não seguia uma sequência incremental fiável. Assim, a *datahora_id* permitiu identificar com precisão qual o último registo inserido na tabela de factos, o que facilitou o controlo de incrementos no carregamento.
4. **Merge Join Causa:** junção entre os dados dos incêndios e a dimensão *Causa*, com base no código de causa, previamente ordenado com os componentes *Sort CodCausa* e *Sort cod_causa*.
5. **Merge Join Local:** junção com a dimensão *Local* através do identificador do local, recorrendo aos módulos *Sort Local* e *Sort local id* para assegurar a ordenação necessária.
6. **Ordenação Final e Carregamento:** após a junção dos dados com as dimensões, foi realizada uma ordenação pela coluna de identificação (*Sort Id*) antes do carregamento final na tabela de destino com o módulo *OLE DB Destination*.

Este fluxo assegurou a consistência das chaves de substituição e permitiu o correto relacionamento entre os dados factuais e as dimensões. A representação visual completa deste fluxo encontra-se disponível no Anexo M – Data Flow tabela de Factos.

F. Estrutura do Dataset Final

A construção do *Data Warehouse* resultou na criação de um dataset final estruturado segundo um modelo dimensional, composto por várias tabelas de dimensões e uma tabela de factos. Cada tabela foi definida com base nos requisitos do negócio e nas transformações aplicadas ao longo do processo ETL.

A tabela no Anexo N apresenta a descrição dos campos que compõem o dataset final, incluindo o nome do campo, o tipo de dados e as principais características associadas a cada atributo. Esta informação foi organizada de forma a permitir uma consulta clara e estruturada dos elementos que integram cada dimensão e a tabela de factos.

Através desta organização, garante-se a consistência, a integridade e a clareza na utilização dos dados, fundamentais para suportar análises e extração de informação relevante.

X. ANÁLISE E NARRATIVA DOS DADOS

Após a construção do *Data Warehouse* e o carregamento dos dados transformados e integrados nas tabelas dimensionais e factuais, foram desenvolvidos três *dashboards* interativos com o objetivo de responder às necessidades de análise dos diferentes utilizadores identificados nas entrevistas e nas *user stories* associadas. Estes *dashboards* permitem uma exploração dinâmica dos dados históricos sobre incêndios em Portugal, em que promove uma melhor compreensão dos fenómenos ao longo do tempo, por região e por causa.

A. Dashboard – Incêndios

O *dashboard* Incêndios (Anexo O) apresentou uma visão geral da evolução dos incêndios em Portugal ao longo dos anos. Este painel permitiu observar padrões temporais de ocorrência, duração e impacto, através de indicadores

relacionados com o número de incêndios, área ardida, duração média e período do dia em que ocorreram.

O *dashboard* integrou os seguintes componentes:

- Filtro temporal – Um *slicer* que permitiu selecionar o intervalo de tempo a analisar (ano, trimestre, mês ou dia).
- Cartão com o número total de incêndios – Apresentou o número total de incêndios registados.
- Cartão com o número de incêndios superiores a 24 horas – Mostrou o total de ocorrências com duração superior a 24 horas.
- Filtro Incêndio superior a 24 horas – Possibilitou isolar os incêndios com duração superior a um dia.
- Gráfico de barras verticais com o número de incêndios por ano – Representou visualmente a evolução anual do número de incêndios.
- Gráfico de barras verticais com o número de incêndios por período do dia – Indicou os períodos com maior frequência de ignições (madrugada, manhã, tarde e noite).
- Cartão com a duração média dos incêndios (horas) – Indicou o valor médio da duração das ocorrências.
- Cartão com a área total ardida (hectares) – Apresentou o total de área ardida no período em análise.
- Gráfico de colunas empilhadas com a área ardida por ano e por tipo de área (agrícola, mato e povoada) – Permitiu analisar a distribuição da área ardida por categoria ao longo dos anos.

Este *dashboard* respondeu aos seguintes requisitos funcionais:

RF01 – Visualizar a área total ardida por ano e por mês

RF03 – Visualizar o número de incêndios por ano

RF06 – Identificar os períodos do dia com mais incêndios

RF07 – Visualizar a duração média dos incêndios

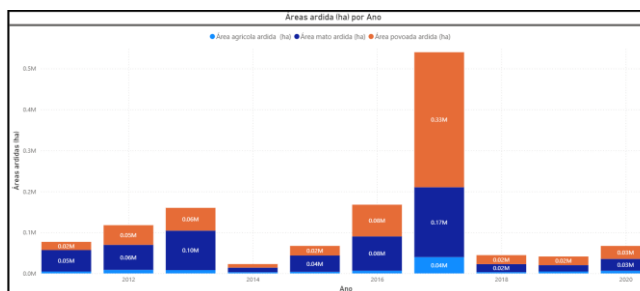
RF08 – Visualizar os incêndios com duração superior a 24 horas

Além disso, permitiu dar resposta às seguintes perguntas definidas na fase de análise de requisitos:

1. Qual foi a área total ardida em cada ano?

Resposta:

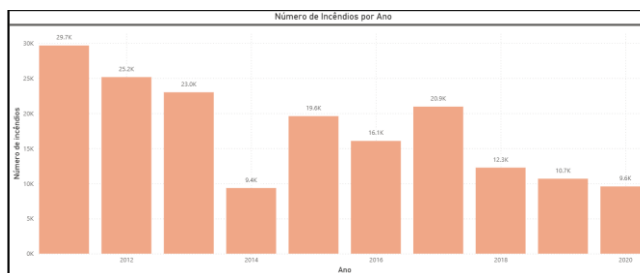
- 2011 – 77.00 K
- 2012 – 117.70 K
- 2013 – 160.01 K
- 2014 – 22.75 K
- 2015 – 67.16 K
- 2016 – 167.71 K
- 2017 – 539.86 K
- 2018 – 44.55 K
- 2019 – 41.30 K
- 2020 – 67.09 K



3. Quantos incêndios ocorreram por ano?

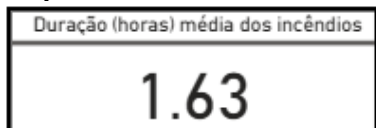
Resposta:

- 2011 – 29658
- 2012 – 25155
- 2013 – 22991
- 2014 – 9351
- 2015 - 19594
- 2016 – 16060
- 2017 – 20948
- 2018 – 12257
- 2019 – 10678
- 2020 – 9579



5. Qual foi a duração média dos incêndios?

Resposta: 1.63 horas.



7. Em que altura do dia ocorreram mais incêndios?

Resposta: Durante o período da tarde.



9. Quantas ocorrências apresentaram uma duração superior a 24 horas?

Resposta: 877 Incêndios.



Para uma melhor compreensão da informação apresentada, recomenda-se a consulta do Anexo O – Dashboard Incêndios, onde se encontraram todos os elementos referidos e a respetiva representação gráfica dos dados.

B. Dashboard – Incêndios por Distrito

O *dashboard* “Incêndios por Distrito” (Anexo P) tem como objetivo apresentar uma visão geográfica da distribuição e impacto dos incêndios em Portugal, permitindo identificar padrões de ocorrência por distrito e concelho.

O *dashboard* é composto pelos seguintes componentes:

- Filtro temporal – Um slicer que permite selecionar o intervalo temporal a analisar (ano, trimestre, mês e dia).
- Cartão com o número total de incêndios – Apresenta o total absoluto de ocorrências registadas.
- Cartão com a área total ardida (ha) – Exibe a soma das áreas ardidas (povoada, agrícola e mato) em hectares.
- Tabela de incêndios por localidade – Contém a informação por distrito e concelho, número de incêndios e área total ardida. Inclui ícones visuais que ajudam a destacar valores relevantes.
- Mapa com o número de incêndios por distrito – Representação geográfica com pontos que indicam a concentração de incêndios.
- Mapa com a área ardida por distrito – Semelhante ao anterior, mas com a dimensão dos pontos proporcional à área ardida.
- Gráfico de barras horizontais com o número de incêndios por distrito e concelho – Compara o número de incêndios entre diferentes localidades.
- Gráfico de barras horizontais com a área ardida por distrito e concelho – Permite comparar a dimensão das áreas ardidas entre distritos e concelhos, em hectares.

Este *dashboard* respondeu aos seguintes requisitos funcionais:

RF05 – Visualizar as zonas do país com mais incêndios

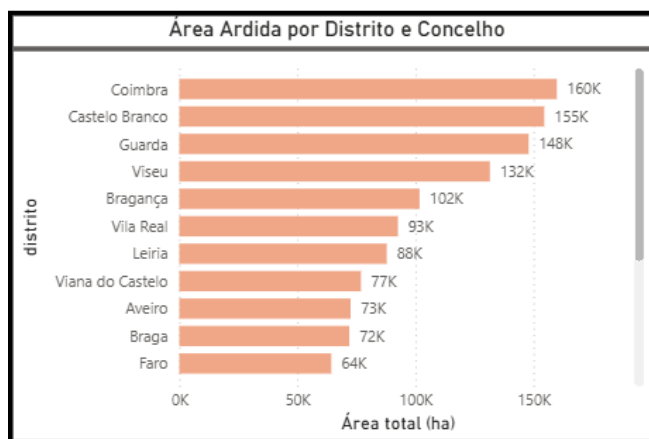
RF12 – Visualizar os distritos e concelhos com maior área ardida

Além disso, permitiu também responder a algumas perguntas definidas na fase de análise de requisitos:

4. Qual é a área ardida por distrito?

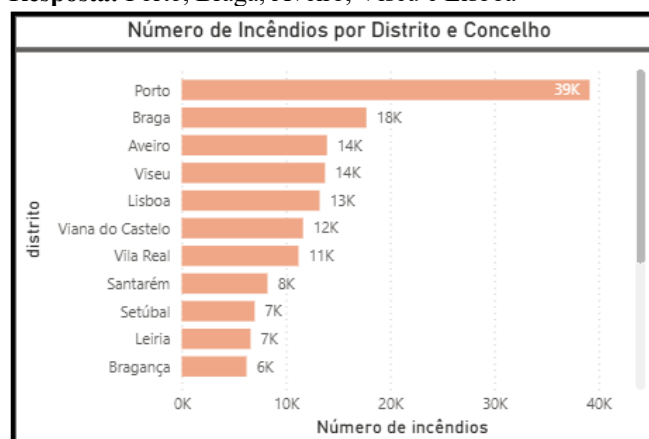
Resposta:

- Coimbra: 160K ha
- Castelo Branco: 155K ha
- Guarda: 148K ha
- Viseu: 132K ha
- Bragança: 102K ha



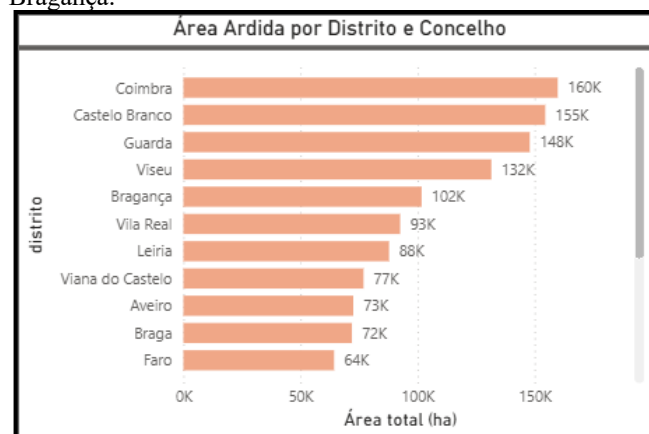
6. Quais são os distritos com mais incêndios?

Resposta: Porto, Braga, Aveiro, Viseu e Lisboa



12. Quais são as cidades ou distritos com maior área ardida?

Resposta: Coimbra, Castelo Branco, Guarda, Viseu e Bragança.



Para uma melhor compreensão da análise e validação das informações apresentadas, recomenda-se a consulta do Anexo P – Dashboard Incêndios por Distrito, onde é possível visualizar em detalhe todos os componentes mencionados e a forma como os dados foram organizados e representados graficamente.

C. Dashboard – Incêndios por Causas

O dashboard “Incêndios por Causa” (Anexo Q) tem como objetivo analisar o número de incêndios registados segundo a

classificação da causa, permitindo identificar os principais fatores de risco e compreender quais os comportamentos, negligências ou intenções mais frequentemente associados às ignições em Portugal.

O mesmo é composto pelos seguintes componentes:

1. Filtro temporal – Que permite selecionar um determinado ano da base de dados para analisar.
2. Filtro por distrito – Permite ao utilizador analisar as causas de incêndios por distrito e por concelho.
3. Tabela com o número de incêndios por causa – Apresenta os campos “Tipo de Causa”, “Grupo Causa” e o número absoluto de incêndios associados a cada uma. Esta ainda inclui uma formatação condicional com barras de cor para destacar os valores mais elevados.

Este painel é particularmente útil para as entidades de prevenção e sensibilização, já que permite identificar quais os comportamentos humanos ou fenómenos que mais frequentemente dão origem a incêndios. A análise regional e temporal permite, por exemplo, verificar se há uma predominância de causas negligentes, intencionais ou desconhecidas em determinadas zonas ou períodos do ano.

Este *dashboard* implementou os requisitos funcionais:

- **RF02** – Visualizar o número de incêndios por causa
- **RF11** – Visualizar a principal causa de incêndios no ano com mais ocorrências

Além disso, permitiu também responder às perguntas estabelecidas inicialmente, sendo as mesmas:

2. Qual é o número de incêndios por causa?

Resposta: Há 5 tipos distintos de causas e 21 grupos de causas. No entanto, os tipos de causas mais frequentes dizem respeito à classe “Desconhecida” e à classe “Intencional” como se pode ver na figura seguinte:

Tipo Causa	Nº de incêndios
Desconhecida	83627
Intencional	27364
Natural	1038
Negligente	49015
Reacendimento	15227
Total	176271

No entanto, estas causas podem ainda ser exploradas em mais detalhe dentro de cada tipo de causa, nomeadamente é possível analisar os seguintes grupos:

tipo_causa	grupo_causa
Desconhecida	Desconhecida
Desconhecida	Indeterminadas
Intencional	Estruturais - Caça e vida selvagem
Intencional	Estruturais - Outras
Intencional	Estruturais - Uso do solo
Intencional	Incendiarismo - Imputáveis
Intencional	Incendiarismo - Inimputáveis
Intencional	Incendiarismo - Sem motivação conhecida
Natural	Naturais
Negligente	Acidentais - Maquinaria
Negligente	Acidentais - Outros
Negligente	Acidentais - Transportes e Comunicações
Negligente	Queimadas de sobrados florestais ou agrícolas
Negligente	Queimadas para gestão de pasto para gado
Negligente	Queimas amontoados de sobrados florestais ou agrícolas
Negligente	Uso do fogo - Fogueiras
Negligente	Uso do fogo - Fumar
Negligente	Uso do fogo - Lançamento Foguetes
Negligente	Uso do fogo - Outros
Negligente	Uso do fogo - Queima de lixo
Reacendimento	Reacendimentos

11. Qual foi a principal causa de incêndios no ano com maior número de ocorrências?

Resposta: Com base na informação de *dashboards* anteriores sabemos que o ano da base de dados em questão em que ocorreram mais incêndios foi no ano de 2011. Dado isto, é possível selecionar esse ano no filtro do período temporal e analisar as maiores causas dos mesmos, que foram uma vez mais o tipo de causa “desconhecida” seguida do tipo de causa “Reacendimentos”.

Além disso, este *dashboard* permite obter insights adicionais, tais como:

- Avaliar a variação das causas ao longo do tempo, aplicando filtros de ano.
- Comparar diferentes distritos quanto às causas predominantes.
- Analisar o peso relativo das causas “Desconhecida” e “Indeterminada”, revelando limitações nos registos ou necessidade de melhoria na investigação de origem.

XI. CONCLUSÃO

O presente projeto permitiu aplicar, de forma prática, os conceitos e metodologias de *Business Intelligence* ao domínio dos incêndios florestais em Portugal. Através da construção de um *data warehouse* e da criação de *dashboards* interativos no Power BI, foi possível transformar dados brutos do Sistema de Gestão de Informação de Incêndios Florestais (SGIF) em informação útil e acessível a diferentes perfis de utilizadores.

Foram cumpridas todas as etapas do ciclo de vida de um projeto de BI desde a recolha e preparação dos dados, modelação dimensional, definição de requisitos, identificação

de utilizadores-alvo, até à construção e validação de *dashboards* com base em KPIs e perguntas orientadoras. Esta abordagem permitiu não só responder às questões previamente definidas, como também gerar insights adicionais sobre a frequência, causas e impacto dos incêndios ao longo do tempo e nas diferentes regiões do país.

A análise realizada evidenciou padrões temporais críticos, como os meses de verão com maior número de ocorrências, bem como causas recorrentes como negligência e reacendimentos. A distribuição geográfica permitiu ainda identificar distritos e concelhos com maior incidência e área ardida, o que poderá apoiar estratégias de prevenção mais direcionadas.

Com este trabalho, reforça-se a importância da análise de dados como suporte à tomada de decisão no setor público, contribuindo para melhorar o planeamento e a eficácia das políticas de prevenção e combate aos incêndios florestais. Ao mesmo tempo, esta experiência revelou-se uma oportunidade de consolidação de competências em BI, desde a construção técnica do sistema até à interpretação estratégica dos dados analisados.

XII. REFERÊNCIAS

- [1] ICNF – Instituto da Conservação da Natureza e das Florestas. *Estatísticas de Fogos Rurais*. Disponível em: <https://www.icnf.pt/florestas/gfr/gfgrgestaoinformacao/estatisticas>. Acedido em: 04/05/2025.
- [2] Wikipedia, *Reorganização Administrativa do Território das Freguesias de 2013*. Disponível em: https://pt.wikipedia.org/wiki/Reorganiza%C3%A7%C3%A3o_Administrativa_do_Territ%C3%B3rio_das_Freguesias_de_2013. Acesso em: 29/06/2025.
- [3] Microsoft. SQL Server Integration Services (SSIS) Documentation. Disponível em: <https://learn.microsoft.com/en-us/sql/integration-services/sql-server-integration-services>. Acedido em: 29/06/2025.
- [4] Microsoft. Visual Studio 2022 Documentation. Disponível em: <https://learn.microsoft.com/en-us/visualstudio>. Acedido em: 29/06/2025.
- [5] Microsoft. Power BI Documentation. Disponível em: <https://learn.microsoft.com/en-us/power-bi>. Acedido em: 30/06/2025.
- [6] Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd Edition). John Wiley & Sons.
- [7] Fi Group. O que são NUTS? Disponível em: <https://pt.fi-group.com/o-que-sao-nuts/>. Acedido em: 30/06/2025.

ANEXOS

Anexo A – Estrutura da Dimensão Data

Nome da Dimensão: **Data**

Tipo Descrição Atributos Importantes	Dimensão			Tabela/Atributo Original			
	Nome	Descrição	Muda?	Tabela	Atributo	Relacionamento	Regras/Domínio
23041978 23/04/1978	data_key						
	data	data (dia/mês/ano)	Não	data			
	dia_da_semana	dia da semana (1 a 7)	Não	data			
	dia_semana_extenso	dia da semana por extenso (segunda-feira a domingo)	Não	data			
	semana	número da semana do ano (1 a ~54)	Não	data			
	mes	mês (1 a 12)	Não	data			
	mes_extenso	mês por extenso (janeiro a dezembro)	Não	data			
	trimestre	trimestre (1 a 4)	Não	data			
	trimestre_extenso	trimestre por extenso (1 a 4 Trimestre)	Não	data			
	semestre	semestre (1 a 2)	Não	data			
	semestre_extenso	semestre por extenso (1 a 2 Semestre)	Não	data			
	ano	Ano (yyyy)	Não	data			
	diaUtil	dia útil da semana	Não	data			Sim, quando o dia da semana for de 2 a 6. Não caso seja 1 ou 7.

Anexo B – Estrutura da Dimensão Hora

Nome da Dimensão: **Hora**

Tipo Descrição Atributos Importantes	Dimensão			Tabela/Atributo Original			
	Nome	Descrição	Muda?	Tabela	Atributo	Relacionamento	Regras/Domínio
1723 17:23 17 23	hora_minutos_key						
	horas_minutos	horas e minutos de um dia	Não	hora			horas: minutos
	horas	horas do dia (24 horas)	Não	hora			
	minutos	minutos da hora	Não	hora			
	periodo_do_dia	período do dia (madrugada / manhã / tarde / noite)	Não	hora			Madrugada SE hora >= 0h E hora < 6h Manhã SE hora >= 6h E < 12h Tarde SE hora >= 12h E hora < 19h Noite SE hora >= 19 E hora < 0h

Anexo C – Estrutura da Dimensão Local

Nome da Dimensão: **Local**

Tipo Descrição Atributos Importantes	Dimensão			Tabela/Atributo Original			
	Nome	Descrição	Muda?	Tabela	Atributo	Relacionamento	Regras/Domínio
	id_local_key	chave artificial					
	DTCCFR	Código INE da freguesia de início do incêndio	Não	local			
	Distrito	Distrito do ponto de início do incêndio	Não	local			
	Concelho	Concelho do ponto de início do incêndio	Não	local			
	Freguesia	Freguesia do ponto de início do incêndio	Sim	local			
	Regiao_Nutsl	Freguesia do ponto de início do incêndio	Não	local			Se distrito = ("Braga", "Bragança", "Porto", "Viana do Castelo", "Vila Real") então regiao_nuts_ii = "Norte" senão se distrito = ("Aveiro", "Castelo Branco", "Coimbra", "Guarda", "Leiria", "Viseu") então regiao_nuts_ii = "Centro" senão se distrito = ("Lisboa", "Setúbal") então regiao_nuts_ii = "Lisboa" senão se distrito = ("Évora", "Beja", "Portalegre", "Santarém") então regiao_nuts_ii = "Alentejo" senão se distrito = "Faro" então regiao_nuts_ii = "Algarve" senão regiao_nuts_ii = "Desconhecido"

Obs.

Regiao_Nutsl é uma coluna derivada com base no distrito.
A coluna Freguesia serão colocadas como "changing attribute" devido às alterações que houve em 2013 nas freguesias
https://pt.wikipedia.org/wiki/Reorganiza%C3%A7%C3%A3o_Administrativa_do_Terr%C3%B3rio_das_Freguesias_de_2013

Anexo D – Estrutura da Dimensão Causa

Nome da Dimensão: **Causa**

Tipo Descrição Atributos Importantes	Dimensão			Tabela/Atributo Original			
	Nome	Descrição	Muda?	Tabela	Atributo	Relacionamento	Regras/Domínio
	id_causa_key	chave artificial					
	CodCausa	Código da causa do incêndio	Não	causa			
	TipoCausa	Tipificação da causa	Não	causa			
	GrupoCausa	Grupo da causa	Sim	causa			
	DescricaoCausa	Descrição da causa	Sim	causa			

Obs.

GrupoCausa e DescricaoCausa podem ser editadas devido a erro humano. Por isso estas duas colunas serão colocadas como "changing attribute"

Anexo E – Estrutura da Tabela de Factos Incêndios

Nome Incendio

Tipo
Descrição
Medidas

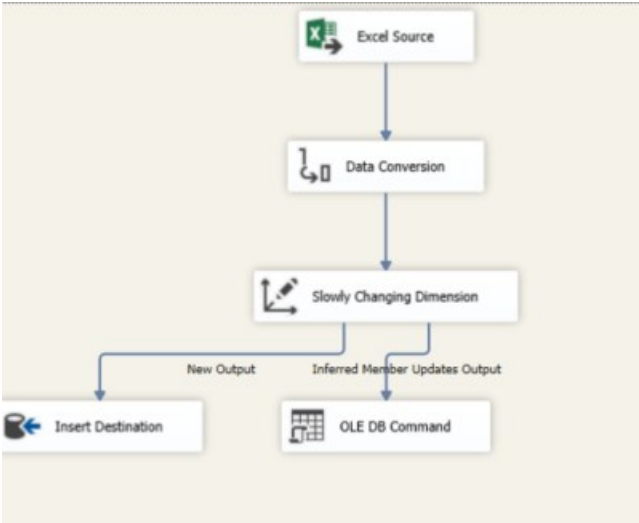
Facto				Origem dos dados			
Nome	Descrição	Formato	Calculado?	Tabela	Atributo	acioname	Regras/Domínio
fact_id				incendios			
Codigo_SGIF	Identificador único SGIF do incêndio rural	Texto	Não	incendios			
AreaPov_ha	Área ardida em povoamentos florestais (ha)	Número decimal com duas casas decimais	Não	incendios			>=0
AreaMato_ha	Área ardida em matos (ha)	Número decimal com duas casas decimais	Não	incendios			>=0
AreaAgric_ha	Área ardida em zonas agrícolas (ha)	Número decimal com duas casas decimais	Não	incendios			>=0
AreaTotal_ha	Área ardida total (povoamentos + matos + zonas agrícolas)	Número decimal com duas casas decimais	Sim	incendios	AreaPov_ha+AreaMato_ha+AreaAgric_ha		>0
ClasseArea	Classe de área ardida total (ha)	Texto	Não	incendios			[0 a 1 ha] SE AreaTotal_ha >0 E <=1 [1 a 10 ha] SE AreaTotal_ha >1 E <=10 [10 a 20 ha] SE AreaTotal_ha >10 E <20 [20 a 50 ha] SE AreaTotal_ha >=20 E <50 [50 a 100 ha] SE AreaTotal_ha >=50 E <100 [100 a 500 ha] SE AreaTotal_ha >=100 E <500 [500 a 1000 ha] SE AreaTotal_ha >=500 E <=1000 [superior a 1000 ha] SE AreaTotal_ha >1000
Duracao_Horas	Duração do incêndio (período entre a data/hora de alerta e a data/hora de extinção)	Número decimal com duas casas decimais	Não	incendios			>=0
IncSup24horas	Incêndio rural com duração superior a 24 horas (período entre a data/hora de alerta e a data/hora de extinção)	Binário	Sim	incendios			1 SE Duracao_Horas >=24.00

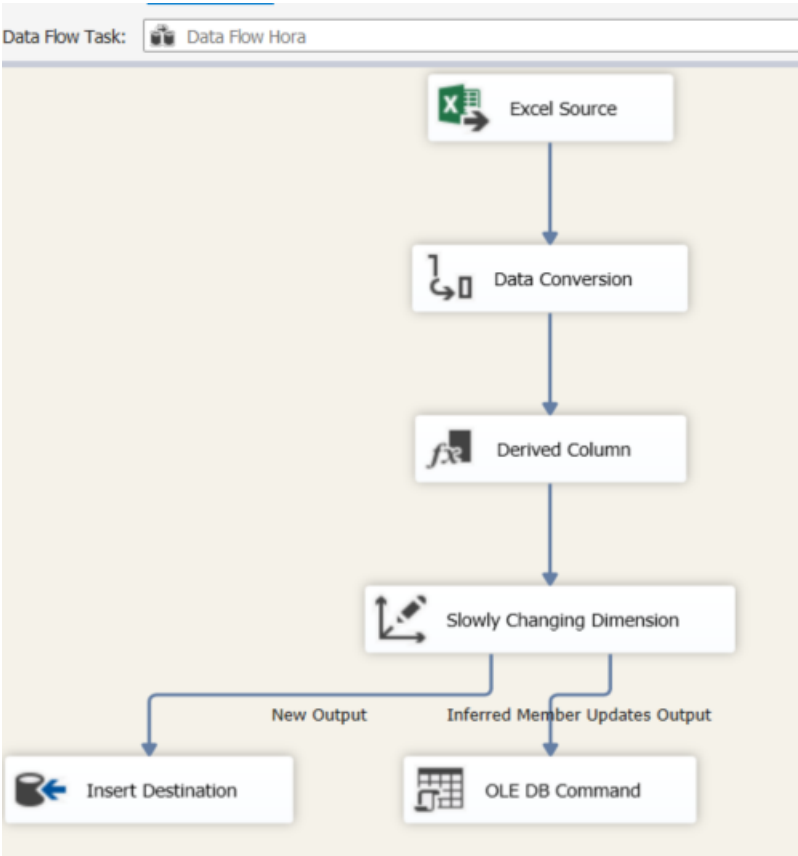
Relacionamentos	Nome	Ligação Facto	Ligação Dimensão
	data_key_inicio		dim_data.data_key
	data_key_fim		dim_data.data_key
	hora_key_inicio		dim_hora.hora_minutos_key
	hora_key_fim		dim_hora.hora_minutos_key
	id_causa_key		dim_causa.id_causa_key
	id_local_key		dim_local.id_local_key

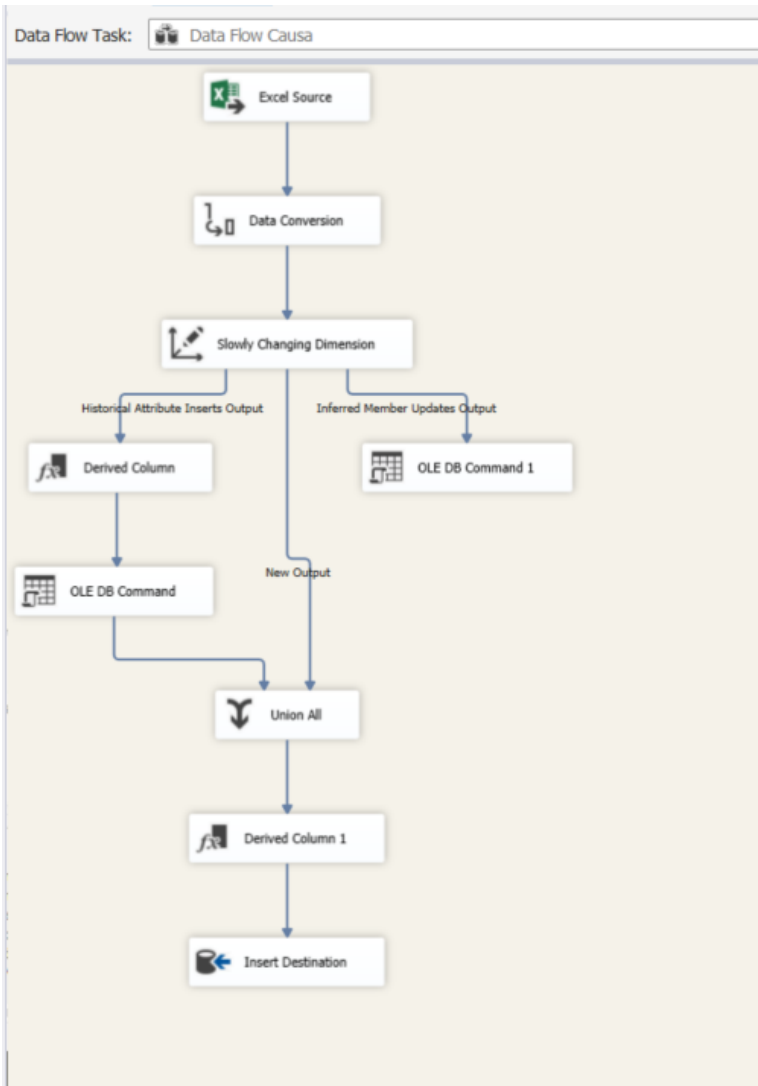
Obs.

O atributo fact_id é calculado com data_key_inicio + data_key_fim + hora_key_inicio + hora_key_fim

Anexo F - Data Flow Dimensão Data







Anexo H.1 – Slow Changing Dimension da Dimensão Causa

Dimension Columns	Change Type
descricao_causa	Historical attribute
grupo_causa	Historical attribute
tipo_causa	Fixed attribute

Anexo I – Carregamento Incremental na Dimensão Causa

Figura I.1 – Número total de registos antes do carregamento

```
select * from [dw_incendios].[dbo].causa where cod_causa = 11;
```

	id_causa_key	cod_causa	tipo_causa	grupo_causa	descricao_causa	expired
1	6	11	Negligente	Uso do fogo - Queima de lixo	Queima de amontoados de lixo	Current

Figura I.2 – Execução do fluxo de dados

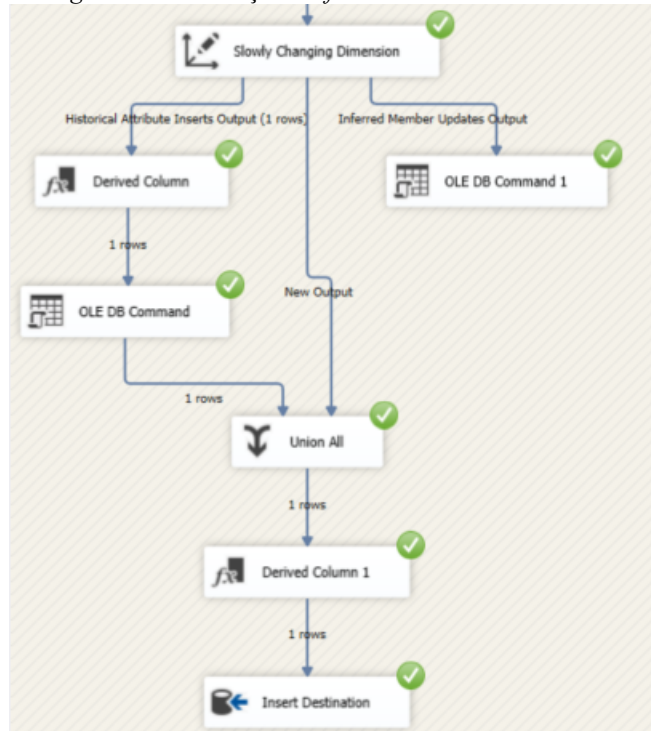
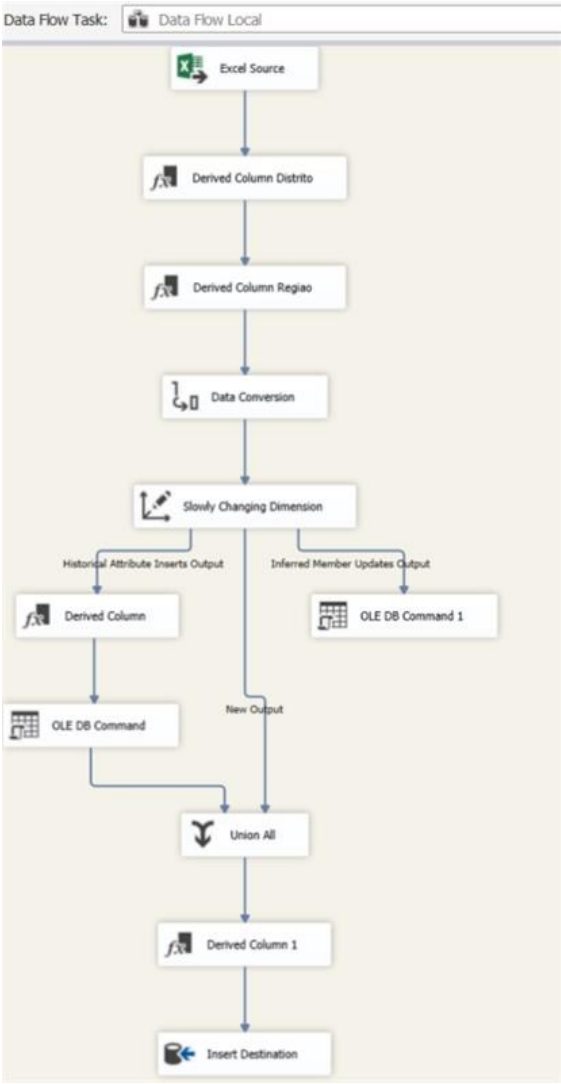


Figura I.3 – Número total de registos após o carregamento

```
select * from [dw_incendios].[dbo].causa where cod_causa = 11;
```

	id_causa_key	cod_causa	tipo_causa	grupo_causa	descricao_causa	expired
1	6	11	Negligente	Uso do fogo - Queima de lixo	Queima de amontoados de lixo	Expired
2	97	11	Negligente	Uso do fogo - Queima de lixo	Queima de amontoados de lixo teste	Current

Anexo J – Data Flow Dimensão Local



Anexo K – Carregamento Incremental na Dimensão Local

Figura K.1 – Número total de registos antes do carregamento

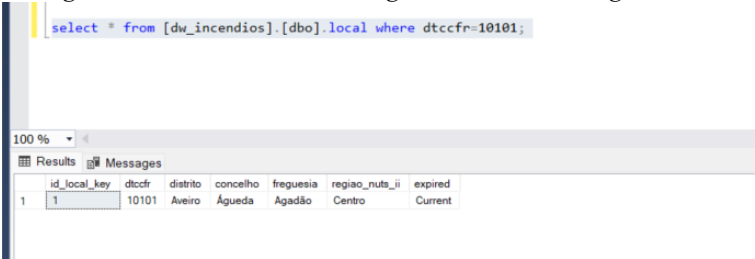


Figura K.2 – Execução do fluxo de dados

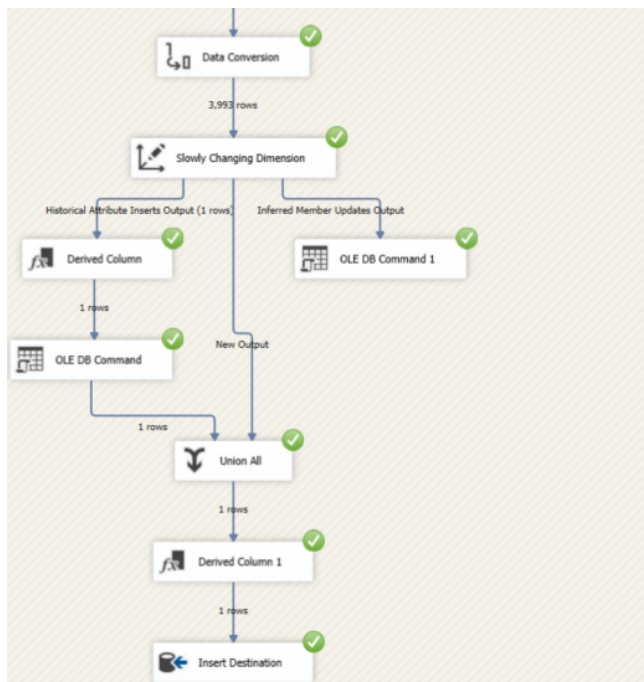


Figura K.3 – Número total de registos após o carregamento

```
select * from [dw_incendios].[dbo].local where dtccfr=10101;
```

	id_local_key	dtccfr	distrito	concelho	freguesia	regiao_nuts_ii	expired
1	1	10101	Aveiro	Águeda	Agadão	Centro	Expired
2	3994	10101	Aveiro	Águeda	Agadão teste	Centro	Current

Anexo L – Carregamento Incremental na tabela de Factos

Figura L.1 – Número total de registos antes do carregamento

```
select count(*) from [dw_incendios].[dbo].fact;
```

	(No column name)
1	176271

Figura L.2 – Execução do fluxo de dados



Figura J.3 – Número total de registros após o carregamento

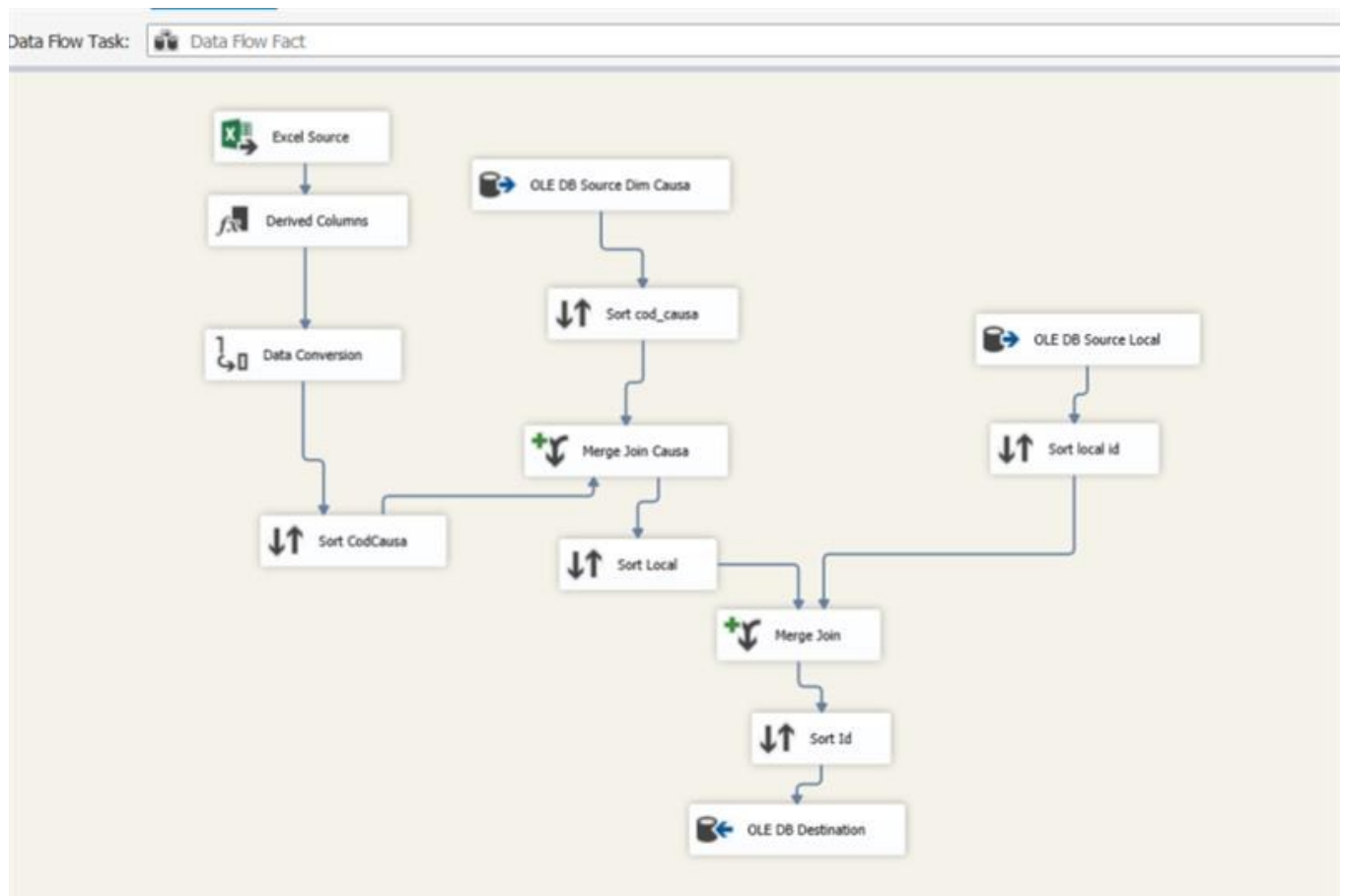
```
select count(*) from [dw_incendios].[dbo].fact;
```

100 %

Results Messages

(No column name)	
1	176272

Anexo M – Data Flow tabela de Factos



Anexo N – Tabelas da estrutura do Dataset Final

Tabela N.1 – Tabela Data

Nome do Campo	Tipo de Dados	Tamanho	Permite Nulos?	Chave Primária?	Descrição
data_key	char	8	Não	Sim	Identificador único da data no formato yyyyymmdd.
data_data	date	-	Não	Não	Representação da data em formato padrão.
data_dia_semana	int	-	Não	Não	Número do dia da semana (1=Domingo, 7=Sábado).
data_dia_semana_extenso	varchar	14	Não	Não	Nome completo do dia da semana (ex: Segunda-feira).
data_semana	int	-	Não	Não	Número da semana no ano (1 a 52).
data_mes	int	-	Não	Não	Número do mês (1 a 12).
data_mes_extenso	varchar	10	Não	Não	Nome do mês por extenso (ex: Janeiro).
data_trimestre	int	-	Não	Não	Número do trimestre (1 a 4).
data_trimestre_extenso	varchar	12	Não	Não	Nome do trimestre (ex: Primeiro Trimestre).
data_ semestre	int	-	Não	Não	Número do semestre (1 ou 2).
data_ semestre_extenso	varchar	10	Não	Não	Nome do semestre por extenso (ex: Segundo Semestre).
data_ano	int	-	Não	Não	Ano correspondente à data.
data_dia_util	char	3	Não	Não	Indicador de dia útil: "Sim" ou "Não".

Tabela N.2 – Tabela Hora

Nome do Campo	Tipo de Dados	Tamanho	Permite Nulos?	Chave Primária?	Descrição
hora_key	char	4	Não	Sim	Identificador único da hora no formato hhmm (ex: 0830 para 08:30).

hora_completa	datetime	-	Não	Não	Representação completa da hora com data e hora (timestamp).
hora	int	-	Não	Não	Hora do dia (0 a 23).
minutos	int	-	Não	Não	Minutos da hora (0 a 59).
periodo_do_dia	varchar	10	Não	Não	Período do dia (ex: “Manhã”, “Tarde”, “Noite”, “Madrugada”).

Tabela N.3 – Tabela Causa

Nome do Campo	Tipo de Dados	Tamanho	Permite Nulos?	Chave Primária?	Descrição
id_causa_key	int	-	Não	Sim	Identificador único da causa no data warehouse.
cod_causa	int	-	Não	Não	Código original da causa proveniente da fonte de dados.
tipo_causa	varchar	15	Não	Não	Tipo da causa do incêndio (ex: Natural, Humana, etc.).
grupo_causa	varchar	60	Não	Não	Agrupamento das causas com base em critérios comuns.
descricao_causa	varchar	100	Não	Não	Descrição detalhada da causa do incêndio.
expired	char	10	Sim	Não	Indicador de validade do registo histórico (ex: “Current”, “Expired”).

Tabela N.4 – Tabela Local

Nome do Campo	Tipo de Dados	Tamanho	Permite Nulos?	Chave Primária?	Descrição
id_local_key	int	-	Não	Sim	Identificador único do local no data warehouse.
dtccfr	int	-	Não	Não	Código INE da freguesia onde

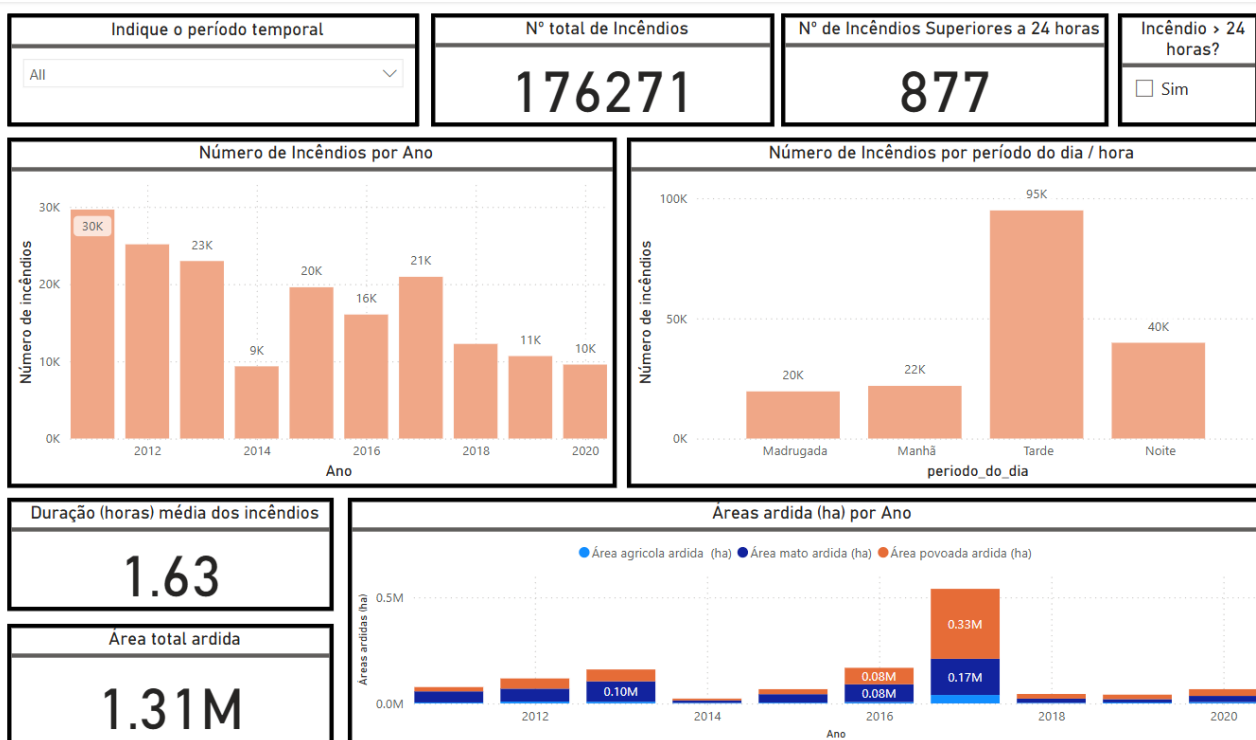
					ocorreu o incêndio.
distrito	varchar	20	Não	Não	Distrito do ponto de ignição do incêndio.
concelho	varchar	30	Não	Não	Concelho do ponto de ignição do incêndio.
freguesia	varchar	60	Não	Não	Freguesia do ponto de ignição do incêndio.
regiao_nuts_ii	varchar	15	Não	Não	Região NUTS II correspondente ao distrito (ex: Norte, Centro, Lisboa, etc.).
expired	char	10	Sim	Não	Indicador de validade do registo histórico (“Current” ou “Expired”).

Tabela N.5 – Tabela de Factos

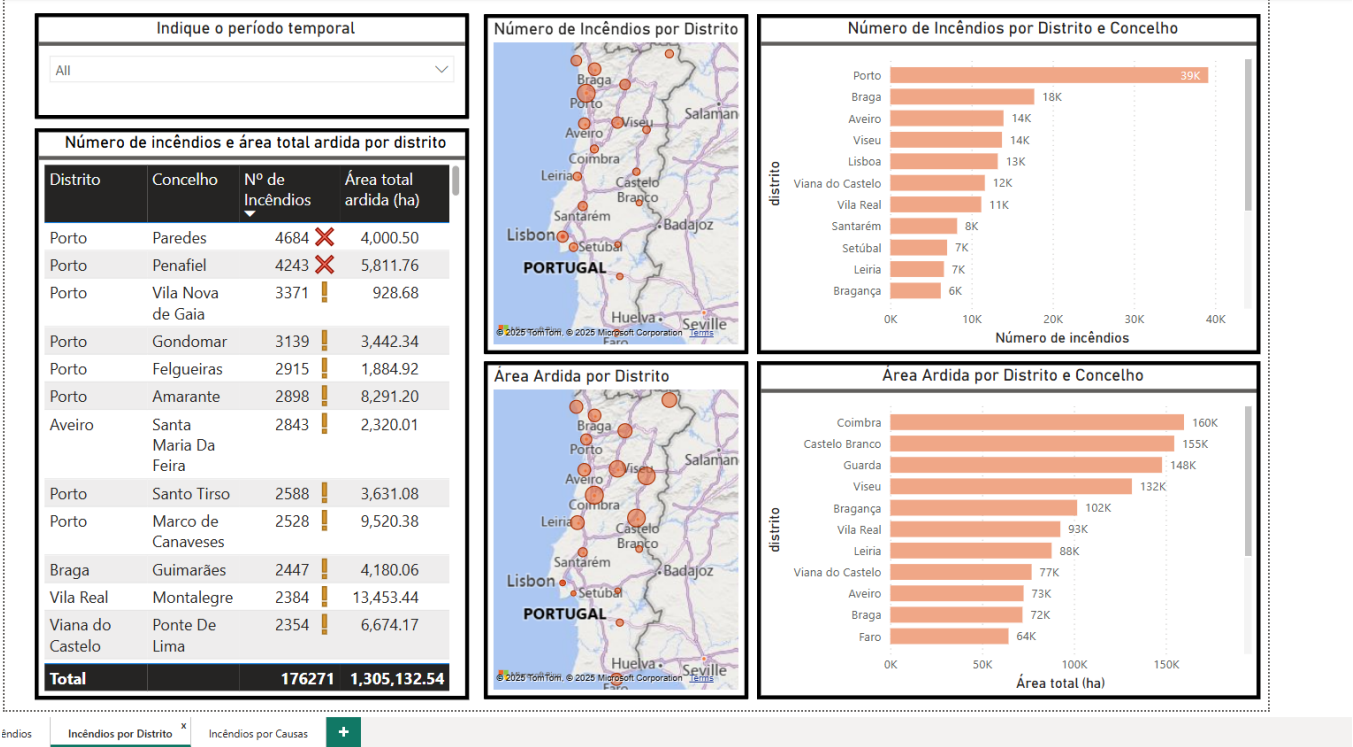
Nome do Campo	Tipo de Dados	Tamanho / Precisão	Permite Nulos?	Chave Primária?	Descrição
fact_id	decimal	24,0	Não	Sim	Identificador único da linha de facto.
codigo_sgif	varchar	15	Não	Não	Código do SGIF atribuído ao incêndio.
data_key_inicio	int	-	Não	Não	Chave da data de início do incêndio.
data_key_fim	int	-	Não	Não	Chave da data de extinção do incêndio.
hora_key_inicio	int	-	Não	Não	Chave da hora de início do incêndio.
hora_key_fim	int	-	Não	Não	Chave da hora de extinção do incêndio.
causa_key	int	-	Não	Não	Chave estrangeira para a dimensão de causas.

local_key	int	-	Não	Não	Chave estrangeira para a dimensão local.
area_pov_ha	decimal	8,2	Não	Não	Área ardida em povoamento florestal (hectares).
area_mato_ha	decimal	8,2	Não	Não	Área ardida em mato (hectares).
area_agric_ha	decimal	8,2	Não	Não	Área ardida em terrenos agrícolas (hectares).
area_total_ha	decimal	8,2	Não	Não	Área total ardida (hectares).
classe_area	varchar	50	Não	Não	Classificação da área total ardida (ex: “Pequeno”, “Médio”, “Grande”).
duracao_horas	float	-	Não	Não	Duração do incêndio em horas.
inc_sup_24_horas	tinyint	-	Não	Não	Indicador binário (0/1) se o incêndio durou mais de 24 horas.

Anexo O – Dashboard Incêndios



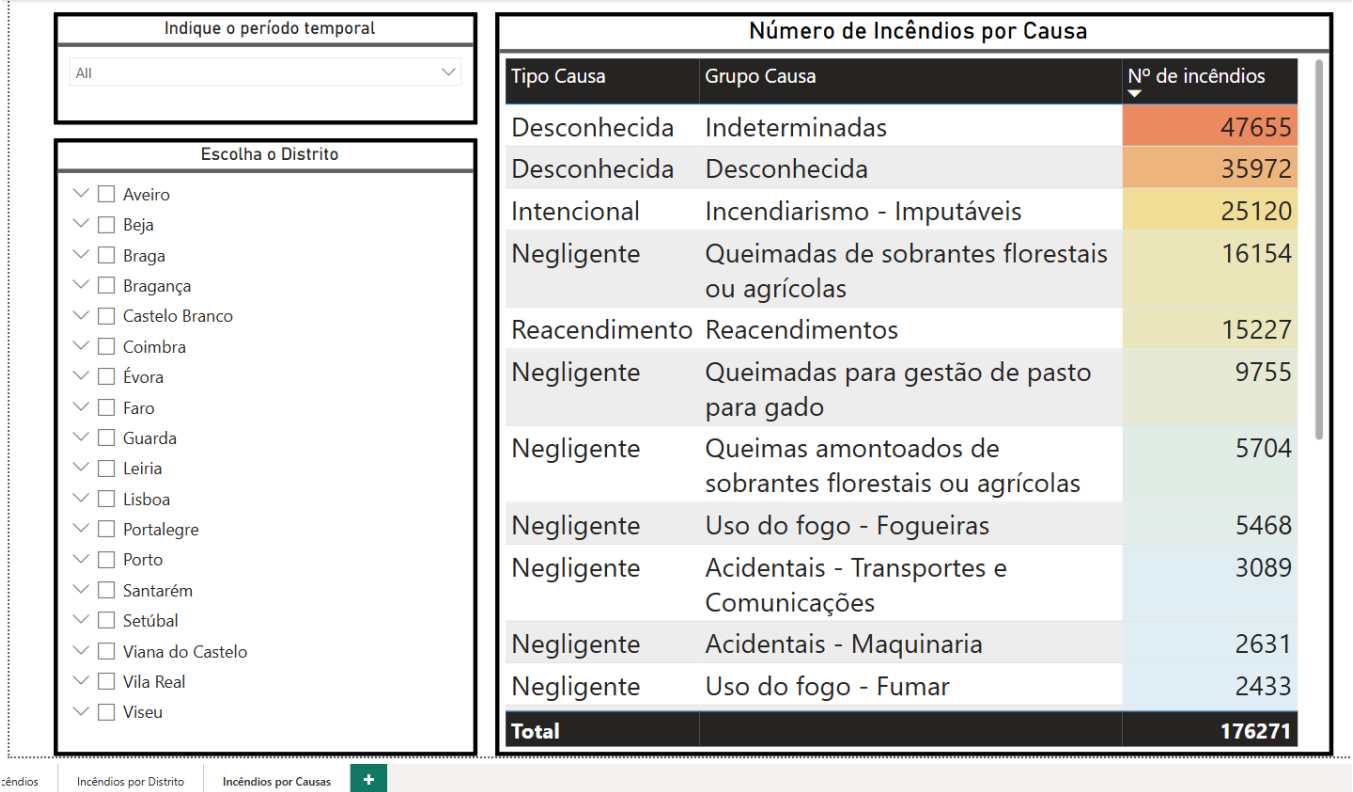
Anexo P – Dashboard Incêndios por Distrito



Incêndios por Distrito

Incêndios por Causas

Anexo Q – Dashboard Incêndios por Causas



Número de Incêndios por Causa

Tipo Causa	Grupo Causa	Nº de incêndios
Desconhecida	Indeterminadas	47655
Desconhecida	Desconhecida	35972
Intencional	Incendiarismo - Imputáveis	25120
Negligente	Queimadas de sobranes florestais ou agrícolas	16154
Reacendimento	Reacendimentos	15227
Negligente	Queimadas para gestão de pasto para gado	9755
Negligente	Queimas amontoados de sobranes florestais ou agrícolas	5704
Negligente	Uso do fogo - Fogueiras	5468
Negligente	Acidentais - Transportes e Comunicações	3089
Negligente	Acidentais - Maquinaria	2631
Negligente	Uso do fogo - Fumar	2433
Total		176271

Incêndios por Distrito

Incêndios por Causas