

# Relatório de Projeto de Business Intelligence - ***Análise de avistamentos de OVNIs***

**Formandos:** Ana Carneiro, Cidália Castro, Johab Santos

**Números:** 8068787, 5579374, 7018655

**UFCD:** 10804 - Business Intelligence

**Data de Entrega:** 28 de Março de 2025





Trabalho realizado no âmbito do curso de Business Intelligence do IEFP, com o objetivo de aplicar técnicas de análise e visualização de dados que nos permitam explorar e detectar padrões em avistamentos de OVNI

s.





## Dedicatória

Dedicamos este projeto a todos os que acreditam no poder do conhecimento, da persistência e do trabalho em equipa.

Àqueles que não desistem perante os desafios e que se superam a cada nova etapa. A nós próprios: Cidália, Ana Carolina e Johab pelo empenho, pela resiliência e pela vontade de evoluir, aprender e fazer sempre melhor.

Este trabalho é fruto de esforço conjunto, mas também de paixão individual por crescer profissionalmente.





## Agradecimentos

Este projeto é o resultado de um verdadeiro trabalho de equipa, desenvolvido em colaboração com os colegas Cidália, Johab e Ana Carolina. Cada um contribuiu com ideias, competências e esforço, e isso fez toda a diferença no resultado final.

Deixamos um especial agradecimento aos formadores que nos acompanharam ao longo da formação:

- Ricardo Pinto Novo, pela orientação nas áreas de ingestão e visualização de dados;
- Dionísio Creoulo, pela partilha de conhecimento sólido em gestão e armazenamento de dados, SQL e normalização;
- Ana Sofia Teixeira, pela abordagem prática e aprofundada à transformação, modelagem e análise avançada de dados;
- Rúben Folha, pelo contributo na componente de storytelling, essencial para comunicar os resultados com clareza e impacto.

Por fim, um agradecimento especial ao IEFP, por proporcionar esta oportunidade de aprendizagem, pelos recursos disponibilizados e pela aposta na qualificação profissional de excelência.







## Epígrafe

*"Os dados podem ser a chave para entender o desconhecido, mas somente quando estamos dispostos a olhar além do convencional."*

Jacques Vallée





## Resumo

Ao longo das eras, o ser humano sempre observou o céu em busca de respostas para os mistérios do universo. Essa curiosidade ancestral continua viva, alimentando questões sobre aquilo que está para além da nossa compreensão. Foi com esse espírito de investigação que se desenvolveu este projeto de Business Intelligence centrado na análise de avistamentos de Objetos Voadores Não Identificados (OVNIs), com base num dataset da plataforma Kaggle.

O trabalho percorreu todas as etapas do ciclo de BI: desde a recolha, limpeza e normalização dos dados numa base relacional em MySQL até à criação de dashboards analíticos no Power BI. Foram aplicadas técnicas de modelação de dados, tratamento de outliers, formatação de colunas e categorização de variáveis. Devido a limitações técnicas na ligação ao servidor SQL, o processo foi adaptado através da importação de ficheiros CSV.

Mais do que visualizar dados, o projeto procurou interpretar e extrair significado da informação: perceber o que os dados comunicam, identificar padrões, cruzar variáveis e levantar hipóteses. Esta abordagem permitiu compreender contextos complexos de forma analítica e quantitativa, traçar relações e, quem sabe, aproximar-nos um pouco mais da verdade... afinal, estará ela mesmo lá fora?

**Palavras-chave:** Business Intelligence, OVNI, UFO, Visualização de Dados, SQL, Power BI, Análise de Dados





## Abstract

Throughout the ages, humans have looked to the skies in search of answers about the universe that surrounds them. This ancestral curiosity continues to inspire questions about what lies beyond our understanding. With that investigative spirit, this Business Intelligence project was developed to analyze Unidentified Flying Object (UFO) sightings, using a dataset from the Kaggle platform.

The project followed all stages of the BI lifecycle: from data collection, cleaning, and normalization in a relational SQL Server database, to the construction of analytical dashboards in Power BI. Techniques such as data modeling, outlier treatment, column formatting, and variable categorization were applied. Due to technical limitations in the direct connection to the SQL server, the team adapted the workflow using CSV file imports.

More than just data visualization, this project aimed to extract meaning from information — identifying patterns, crossing variables, and forming hypotheses. This approach enabled the team to analyze complex contexts quantitatively and uncover relationships that may bring us closer to an answer. After all... is the truth really out there?

**Keywords:** Business Intelligence, UFO, Data Analysis, Data Visualization, Power BI, SQL





# Índice

Dedicatória.....	5
Epígrafe.....	9
Resumo.....	11
Abstract.....	13
Índice.....	15
Índice de Quadros e Figuras.....	18
1. Introdução.....	20
2. Objetivos.....	21
2.1 Objetivo Geral.....	21
2.2 Objetivos Específicos.....	21
3. Recolha e processamento dos dados.....	22
3.1 Pré-Análise em Excel.....	22
3.2 Problemas detectados no dataset original.....	22
4. Criação da Base de Dados e Inserção dos Dados.....	24
4.1 Criação da Base de Dados.....	24
4.2 Inserção dos Dados.....	24
5. Normalização da Base de Dados.....	25
5.1 Atribuição de IDs, Criação de Tabelas e Inserção de Dados.....	25
6. Importação, Limpeza e Transformação de Dados.....	28
6.1 Importação dos Dados no Power BI.....	28
6.2 Relacionamento entre tabelas.....	29
6.3 Transformação e Limpeza.....	29
6.4 Tratamento de Outliers.....	30
6.5 Processos complementares fora do Power Query.....	31
7. Construção do Dashboard.....	34
7.1 Cartões de Indicadores (KPI).....	34
7.2 Top 5 Formas.....	34
7.3 Top 5 Países.....	35
7.4 Evolução Temporal.....	35
7.5 Mapa Geográfico.....	36
7.6 Word Cloud.....	36
7.7 Avistamentos por Categoria de Duração.....	37
7.8 Avistamentos por Dia de Semana.....	38
7.9 Segmentadores Interativos.....	38
8. Dashboard - Visão Geral.....	39
9. Dashboard - Exploração e Detalhe.....	39



---

Conclusão.....	40
Bibliografia e Webgrafia.....	42
Lista de anexos.....	43







## Índice de Quadros e Figuras

Figura 1: Script SQL para criação de coluna e atribuição de IDs únicos.....	26
Figura 2: Script SQL para criação de tabela e inserção de dados.....	27
Figura 3: Relacionamentos entre tabelas no Power BI.....	29
Figura 4: Linguagem M para ligação à API.....	30
Figura 5: Medidas DAX para cálculo do 1º e 3º quartil.....	31
Figura 6: Linguagem M para criação de coluna personalizada e identificação de outliers.....	31
Figura 7: Criação de coluna em DAX para categorização de duração.....	32
Figura 8: Criação de tabela em DAX para calendário.....	32
Figura 9: Medida DAX para cálculo de média de duração em formato legível....	33
Figura 10: TOP 5 formas de OVNI.....	34
Figura 11: TOP 5 Países com Mais Avistamentos.....	35
Figura 12: Evolução de avistamentos ao longo do tempo.....	35
Figura 13: Avistamentos por localização.....	36
Figura 14: Word Cloud.....	37
Figura 15: Avistamentos por categoria de duração.....	37
Figura 16: Avistamentos por dia da semana.....	38
Figura 17: Dashboard - Visão Geral.....	39
Figura 18: Dashboard - Exploração e Detalhe.....	39





## 1. Introdução

O ser humano sempre se fascinou pelo desconhecido. Entre os muitos mistérios que nos rodeiam, os avistamentos de Objetos Voadores Não Identificados (OVNI)s continuam a despertar curiosidade, especulação e interesse científico. A análise de registos de avistamentos é, assim, uma oportunidade não só para explorar padrões de comportamento e distribuição geográfica, mas também para aplicar metodologias modernas de tratamento e visualização de dados.

Neste contexto, o presente projeto de Business Intelligence surge como um desafio académico e técnico, tendo como objetivo aplicar o ciclo completo de BI sobre um dataset real e complexo. Através da utilização de ferramentas como Excel, MySQL e Power BI, procurou-se estruturar e analisar uma base de dados de avistamentos de OVNI)s, recolhida na plataforma Kaggle.

Esta introdução marca o ponto de partida para uma jornada analítica que pretende ir além da simples visualização, com o objetivo de interpretar dados, identificar padrões e lançar hipóteses num dos temas mais enigmáticos e instigantes da atualidade.



## 2. Objetivos

### 2.1 Objetivo Geral

Desenvolver um projeto completo de Business Intelligence aplicado à análise de avistamentos de Objetos Voadores Não Identificados (OVNI)s, utilizando um dataset real da plataforma Kaggle, com o intuito de aplicar, integrar e consolidar competências em recolha, transformação, modelação, análise e visualização de dados.

### 2.2 Objetivos Específicos

- Explorar e compreender a estrutura de um dataset real e semi-estruturado;
- Realizar a limpeza e normalização dos dados numa base relacional criada em SQL Server;
- Aplicar boas práticas de modelação de dados, assegurando integridade e eficiência nas relações entre tabelas;
- Importar, transformar e preparar os dados no Power BI, corrigindo formatos e tratando outliers;
- Criar colunas calculadas e medidas DAX para análises avançadas;
- Desenvolver dashboards interativos que permitam identificar padrões geográficos, temporais e de comportamento nos avistamentos;
- Estimular o raciocínio analítico e crítico perante dados complexos, promovendo a tomada de decisões baseadas em evidência.



### 3. Recolha e processamento dos dados

A primeira etapa do projeto consistiu na seleção e compreensão do dataset a ser trabalhado. A fonte de dados escolhida foi a plataforma Kaggle, conhecida pela diversidade e qualidade dos seus conjuntos de dados públicos. A pesquisa foi orientada por interesse temático (avistamentos de OVNI)s e pela potencial complexidade dos dados disponíveis, o que nos permitiria aplicar todas as fases do ciclo de Business Intelligence.

O dataset selecionado continha dois ficheiros principais: `complete.csv` e `scrubbed.csv`. Optámos por utilizar o `scrubbed.csv`, uma vez que apresentava uma versão mais limpa, sem registos incompletos ou inválidos, facilitando o início do tratamento e modelação.

#### 3.1 Pré-Análise em Excel

Antes da criação da base de dados, realizámos uma análise exploratória inicial no Microsoft Excel, com os seguintes objetivos:

- **Remoção de duplicados:** Identificámos e eliminamos linhas repetidas com base em combinações de colunas como data, localização e descrição.
- **Criação de um identificador único (ID):** Atribuímos um ID a cada avistamento para facilitar a indexação e posterior normalização dos dados.
- **Avaliação das colunas disponíveis:** Identificámos campos que apresentavam elevada taxa de valores nulos, nomes inconsistentes ou formatos desajustados.

#### 3.2 Problemas detectados no dataset original

Durante esta fase, foram identificados diversos problemas que necessitavam de intervenção:



- Nomes de colunas com espaços e caracteres especiais, o que dificulta a utilização em SQL e Power BI. Exemplos: Duration (seconds) e Date Posted;
- A coluna DateTime teve de ser renomeada, pois entrava em conflito com a palavra reservada do SQL;
- Formatos inconsistentes em colunas de data e duração, com entradas ambíguas ou em formato de texto;
- Campos com baixa relevância ou pouca qualidade, como:
  - Posted: data de publicação do avistamento, sem utilidade analítica direta;
  - City e Country: com muitos valores nulos, optando por manter a longitude e latitude e posteriormente renomear a coluna com base nestes dados mais específicos.

Esta primeira análise permitiu estabelecer as bases para a criação da base de dados relacional, garantindo que apenas dados relevantes e em condições mínimas de qualidade fossem mantidos para as fases seguintes do projeto.



## 4. Criação da Base de Dados e Inserção dos Dados

Após o tratamento preliminar dos dados em Excel, iniciámos a construção da base de dados relacional utilizando o MySQL Workbench, com o objetivo de organizar a informação de forma estruturada, facilitar o processo de normalização e permitir consultas eficientes.

### 4.1 Criação da Base de Dados

Foi criada uma base de dados com o nome *area52*, em alusão à conhecida “Área 51”, reforçando a identidade do projeto e a sua ligação ao tema dos OVNIs. A criação da base de dados foi feita através do ambiente gráfico do MySQL Workbench, utilizando comandos SQL e ferramentas disponíveis na plataforma.

### 4.2 Inserção dos Dados

A tabela principal com os dados brutos foi importada diretamente a partir do ficheiro previamente tratado *area52.csv* através do Table Import Wizard, uma ferramenta integrada do MySQL Workbench que permite importar ficheiros CSV de forma rápida e intuitiva para tabelas já criadas.





## 5. Normalização da Base de Dados

Após a importação dos dados brutos para o MySQL Workbench, avançamos para a fase de normalização, com o objetivo de garantir uma estrutura de base de dados eficiente, sem redundâncias, e que permitisse análises mais robustas no Power BI.

A normalização seguiu os princípios clássicos até à 3ª Forma Normal, assegurando que cada tabela representasse uma entidade distinta, com relações bem definidas e integridade referencial mantida através de chaves primárias e estrangeiras.

A partir da tabela original area52, foram identificadas três entidades principais e que justificavam a separação em tabelas próprias:

- **Localização (location):** latitude e longitude
- **Forma (shape):** tipo de objeto avistado
- **Duração (duration):** tempo de duração do avistamento

### 5.1 Atribuição de IDs, Criação de Tabelas e Inserção de Dados

Os seguintes passos foram realizados:

1. Adição de colunas de ID na tabela area52, como indicado na *figura 1*
  - a. Foram criadas as colunas location\_id, shape\_id e duration\_id;
  - b. Utilizámos a função ROW\_NUMBER() em subqueries para gerar IDs únicos para cada valor distinto.



```
-- Criar a coluna `location_id`  
ALTER TABLE area52 ADD COLUMN location_id INT;  
  
-- Atribuir IDs únicos a localizações com base em latitude e longitude  
UPDATE area52 a  
JOIN (  
    SELECT latitude, longitude,  
           ROW_NUMBER() OVER (ORDER BY latitude, longitude) AS location_id  
    FROM (SELECT DISTINCT latitude, longitude FROM area52) AS unique_locations  
  ) temp ON a.latitude = temp.latitude AND a.longitude = temp.longitude  
SET a.location_id = temp.location_id;
```

Figura 1: Script SQL para criação de coluna e atribuição de IDs únicos

2. Atribuição de location\_id, como demonstra a *figura 1*
  - a. IDs únicos foram atribuídos com base na combinação latitude + longitude.
3. Atribuição de shape\_id
  - a. Foram atribuídos IDs únicos a cada forma distinta (shape).
4. Atribuição de duration\_id
  - a. A coluna durationhoursmin foi descartada, e os IDs foram atribuídos com base apenas em durationseconds.
5. Criação das Tabelas Normalizadas, como podemos visualizar na *figura 2*
  - a. Foram criadas três tabelas normalizadas:
    - i. locations (id, latitude, longitude)
    - ii. shapes (id, shape\_name)
    - iii. durations (id, durationseconds)\



```
-- Criar a tabela `locations`  
CREATE TABLE locations (  
  id INT PRIMARY KEY AUTO_INCREMENT,  
  latitude DOUBLE NOT NULL,  
  longitude DOUBLE NOT NULL,  
  UNIQUE(latitude, longitude) -- Evita duplicação  
);  
  
INSERT INTO locations (latitude, longitude)  
SELECT DISTINCT latitude, longitude FROM area52;
```

Figura 2: Script SQL para criação de tabela e inserção de dados

## 6. Criação da Tabela Final sightings

- a. Foi então criada a tabela principal sightings, composta pelos seguintes campos:
  - i. id\_sighting (PK)
  - ii. datetime\_sighting (formato texto)
  - iii. location\_id (FK)
  - iv. shape\_id (FK)
  - v. duration\_id (FK)
  - vi. comments (texto)

## 7. Migração dos Dados

- a. Os dados foram transferidos da tabela original area52 para a nova estrutura com base em INSERT INTO e SELECT, como indicado na *figura 2*, respeitando as ligações entre as tabelas. Por fim, foi realizada a validação da migração, com uma query para verificar se existiam valores nulos nos campos de ligação (location\_id, shape\_id, duration\_id);
- b. Após validação, a tabela temporária area52 foi eliminada, concluindo a transição para a estrutura relacional normalizada.



## 6. Importação, Limpeza e Transformação de Dados

Com a base de dados relacional finalizada, avançámos para a fase de ETL (Extract, Transform, Load) no Power BI, onde os dados foram preparados para visualização e análise.

### 6.1 Importação dos Dados no Power BI

A abordagem inicial consistiu em estabelecer uma ligação direta entre o Power BI e a base de dados MySQL através do conector nativo. A ligação foi bem-sucedida e permitiu importar as tabelas necessárias para o ambiente do Power Query.

No entanto, à medida que realizávamos transformações mais complexas ou alterações relevantes no Power Query, a ligação à base de dados tornava-se instável, resultando em erros de tempo de espera e perdas de ligação ao servidor.

Este problema repetido comprometia a continuidade do desenvolvimento do modelo no Power BI e impedia o fluxo de trabalho fluido necessário nesta fase do projeto.

Dado o constrangimento técnico, tomámos a decisão de exportar todas as tabelas normalizadas do MySQL para ficheiros CSV. Esta alternativa garantiu total controlo sobre os dados e permitiu:

- Importar os ficheiros diretamente para o Power BI;
- Trabalhar localmente sem depender da estabilidade da ligação com o servidor;
- Acelerar o processo de transformação e modelação no Power BI.



## 6.2 Relacionamento entre tabelas

Antes de iniciar qualquer transformação ou limpeza, foi garantido que as relações entre as tabelas estavam corretamente definidas no modelo de dados do Power BI, replicando a estrutura relacional criada no MySQL. Estas ligações basearam-se nas chaves estrangeiras:

- sightings[location\_id] → locations[id] - [N - 1]
- sightings[shape\_id] → shapes[id] - [N - 1]
- sightings[duration\_id] → durations[id] - [N - 1]

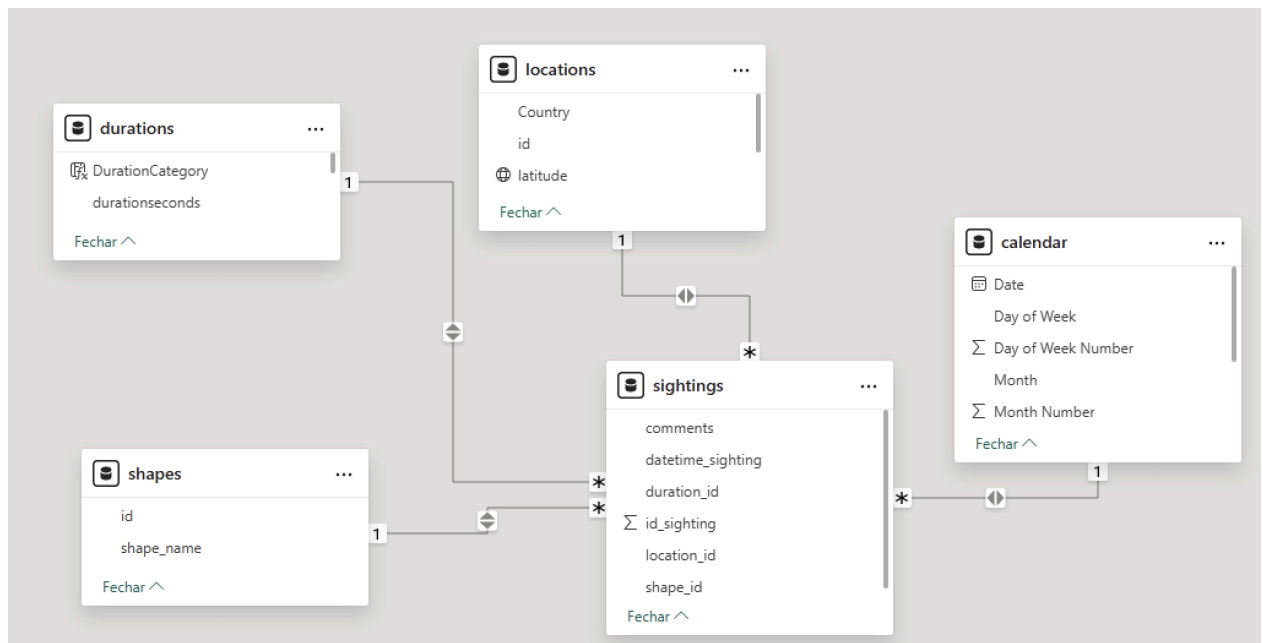


Figura 3: Relacionamentos entre tabelas no Power BI

## 6.3 Transformação e Limpeza

Dentro do Power Query foram realizadas diversas operações de limpeza, enriquecimento e preparação de dados, nomeadamente:

- **Conversão de tipos de dados:** datas, números e textos foram ajustados para os tipos corretos;



- **Tratamento de valores nulos ou inconsistentes:** linhas incompletas ou inválidas foram removidas;
- **Limpeza textual:** foram removidos espaços em excesso, caracteres especiais e padronizados os campos de texto como comments e shape\_name;
- **Identificação geográfica:** para enriquecer a análise espacial, foi feita uma requisição via API a partir das coordenadas (latitude e longitude) para obter o nome do local correspondente, através de uma coluna personalizada, conforme *figura 4*;

```
let
    Latitude = [latitude],
    Longitude = [longitude],
    URL = "https://nominatim.openstreetmap.org/reverse?format=json&lat=" & Text.From(Latitude)
    & "&lon=" & Text.From(Longitude),
    JSON = try Json.Document(Web.Contents(URL)) otherwise null, // Evita erro caso a API não
    retorne resposta
    Address = try JSON[address] otherwise null, // Se não houver "address", retorna null
    City = try Address[city] otherwise try Address[town] otherwise try Address[village] otherwise
    try Address[municipality] otherwise "Unknown city",
    Country = try Address[country] otherwise "Unknown country",
    Cidade_Pais = if City = "Unknown city" and Country = "Unknown country" then "Unknown
    location" else City & ", " & Country
in
    Cidade_Pais
```

*Figura 4: Linguagem M para ligação à API*

Estas transformações asseguraram que os dados estivessem limpos, consistentes e prontos para análise.

## 6.4 Tratamento de Outliers

Um dos aspetos mais críticos da preparação dos dados foi o tratamento de outliers na coluna durationseconds, que apresentava valores extremamente altos e fora do padrão esperado para um avistamento.

Para este fim, foi utilizada uma abordagem estatística baseada no Intervalo Interquartil (IQR), mas implementada diretamente em DAX, com os seguintes passos:



1. Cálculo do 1º Quartil (Q1) e do 3º Quartil (Q3) para durationseconds, com medidas DAX, conforme *figura 5*:

```
Q1_Duration = PERCENTILEX.INC(durations, durations[durationseconds], 0.25)
```

```
Q3_Duration = PERCENTILEX.INC(durations, durations[durationseconds], 0.75)
```

*Figura 5: Medidas DAX para cálculo do 1º e 3º quartil*

2. Criação de uma nova coluna com um filtro lógico que identificava se um registo era ou não um outlier, conforme *figura 6*:

Nome da coluna nova

durationsecondssemoutliers

Fórmula de coluna personalizada ⓘ

```
= let
    Q1 = 203.75,
    Q3 = 3375,
    IQR = Q3 - Q1,
    lower = 0,
    upper = Q3 + 1.5 * IQR
in
    if [durationseconds] >= lower and [durationseconds] <=
    upper then 0 else 1
```

*Figura 6: Linguagem M para criação de coluna personalizada e identificação de outliers*

Após criação desta coluna personalizada, filtramos os valores 0, retirando desta forma os outliers.

## 6.5 Processos complementares fora do Power Query

Após a fase de ETL, outras operações foram realizadas na área de modelação e com o uso de DAX:



- Categorização da duração dos avistamentos com base na coluna `durationseconds`, criando grupos como Very Short, Short, Long e Very Long, conforme *figura 7*;

```

1 DurationCategory =
2 SWITCH(TRUE(),
3     Durations[durationseconds] <= 60, "Very Short",
4     Durations[durationseconds] <= 300, "Short",
5     Durations[durationseconds] <= 1800, "Medium",
6     Durations[durationseconds] <= 7200, "Long",
7     "Very Long"
8 )

```

*Figura 7: Criação de coluna em DAX para categorização de duração*

- Criação de tabela calendário, conforme *figura 8*, para análises temporais detalhadas, com a identificação de padrões sazonais e tendências ao longo do tempo. Este calendário foi relacionado com a tabela sightings através do campo `date`, conforme já demonstrado anteriormente (*figura 3*);

```

calendar =
ADDCOLUMNS (
    CALENDAR (DATE(1949,1,1), DATE(2015,12,31)),
    "Year", YEAR([Date]),
    "Month", FORMAT([Date], "MMMM"),
    "Month Number", MONTH([Date]),
    "Day of Week", FORMAT([Date], "dddd"),
    "Day of Week Number", WEEKDAY([Date]),
    "Quarter", "Q" & FORMAT([Date], "Q"),
    "Year-Month", FORMAT([Date], "YYYY-MM")
)

```

*Figura 8: Criação de tabela em DAX para calendário*





- Criação de medida DAX para cálculo da média de duração de avistamento em formato legível (minutos e segundos), evitando representações confusas como 1.5K, conforme *figura 9*;

```
AverageDurationFormatted =  
VAR TotalSeconds = AVERAGE(Durations[durationseconds])  
VAR Minutes = INT(TotalSeconds / 60)  
VAR Seconds = ROUND(MOD(TotalSeconds, 60), 0)  
RETURN  
Minutes & " min " & Seconds & " sec"
```

*Figura 9: Medida DAX para cálculo de média de duração em formato legível*



## 7. Construção do Dashboard

Com os dados preparados, limpos e organizados num modelo relacional sólido, iniciámos a construção das visualizações no Power BI, com o objetivo de explorar e comunicar insights relevantes de forma clara, interativa e visualmente apelativa.

Foram desenvolvidos vários tipos de visualizações, cada uma com propósitos distintos, desde a apresentação de indicadores globais até à análise geográfica, temporal e categórica dos avistamentos.

### 7.1 Cartões de Indicadores (KPI)

Foram utilizados cartões simples para destacar métricas de alto nível:

- **Total de Avistamentos:** Contagem global dos registos válidos.
- **Duração Média dos Avistamentos:** Criada com medida DAX que converte a média de segundos para um formato legível em minutos e segundos, evitando o uso de valores como 1.5K, que dificultavam a interpretação, conforme *figura 9*.

### 7.2 Top 5 Formas

Foi criado um gráfico de barras horizontais para mostrar os cinco formatos mais frequentemente reportados, conforme *figura 10*. Esta visualização permitiu perceber quais os tipos de objetos mais observados.

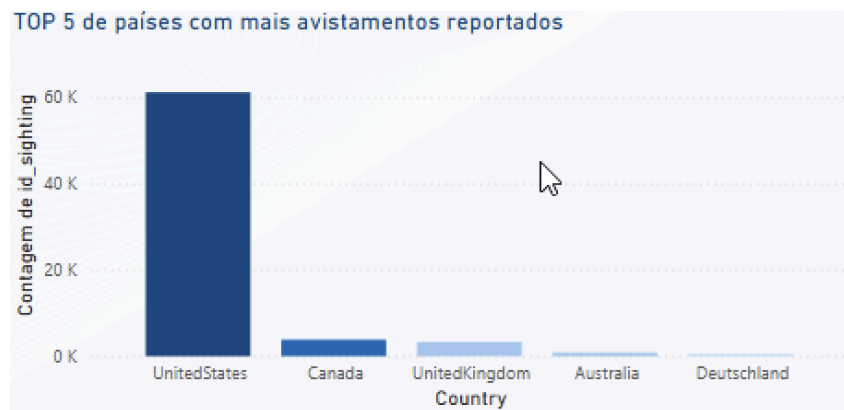
*Figura 10: TOP 5 formas de OVNI's*





### 7.3 Top 5 Países

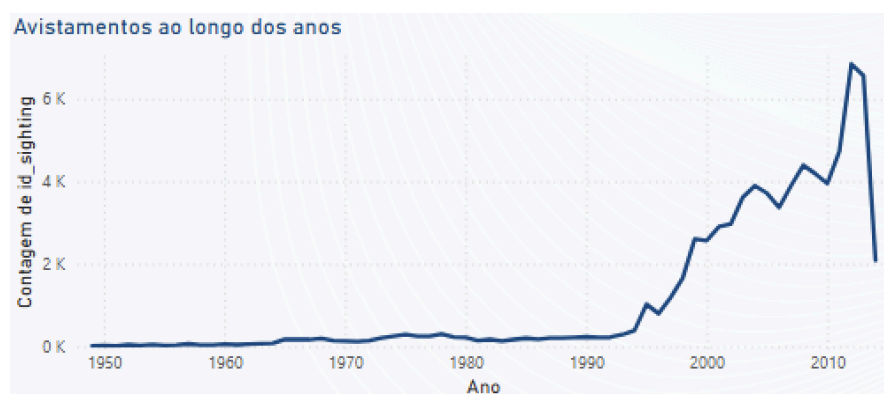
O gráfico de colunas foi construído para destacar os cinco países com maior número de registos, ajudando a identificar zonas geográficas com maior incidência de avistamentos, conforme *figura 11*.



*Figura 11: TOP 5 Países com Mais Avistamentos*

### 7.4 Evolução Temporal

Criado com base na extração do ano da data de avistamento, este gráfico mostrou a evolução anual do número de registos, permitindo observar a tendência temporal, conforme *figura 12*.



*Figura 12: Evolução de avistamentos ao longo do tempo*



## 7.5 Mapa Geográfico

O mapa foi construído com base nos campos de latitude e longitude, possibilitando a visualização geoespacial dos avistamentos. Cada ponto representa um avistamento, com possibilidade de interação e filtragem cruzada com outras visualizações, conforme *figura 13*.



*Figura 13: Avistamentos por localização*

## 7.6 Word Cloud

Com base na coluna comments, previamente limpa no Power Query, foi criada uma word cloud que destacou as palavras mais utilizadas nas descrições dos avistamentos, conforme *figura 14*.

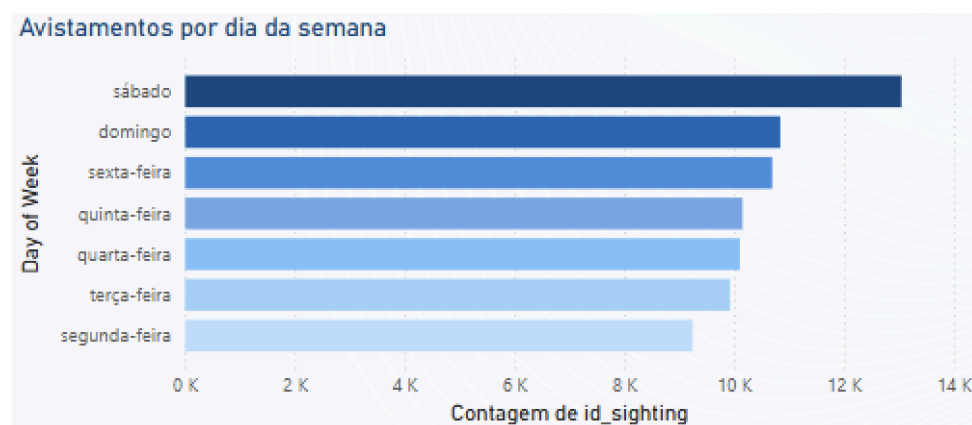
Esta visualização trouxe uma camada qualitativa à análise, evidenciando termos recorrentes como “light”, “disk”, “moving”, entre outros.





## 7.8 Avistamentos por Dia de Semana

Através do dia da semana do calendário criado em DAX, foi construído um gráfico de barras que evidencia em que dias ocorrem mais avistamentos. Esta análise permite identificar possíveis padrões comportamentais ou ambientais, conforme *figura 16*.



*Figura 16: Avistamentos por dia da semana*

## 7.9 Segmentadores Interativos

Foram adicionados segmentadores para permitir ao utilizador filtrar dinamicamente os dados com base em:

- Ano do avistamento
- País
- Tipo de forma
- Categoria da duração

Estes filtros oferecem maior flexibilidade à análise e melhoram a navegação entre as diferentes visualizações do relatório. No entanto, devido à extensão dos dados, foram aplicados filtros aos segmentadores, de forma a aparecerem apenas os 6 anos com mais avistamentos, 5 países com maior número de avistamentos e, do mesmo modo, 5 formas mais avistadas.





## Conclusão

O desenvolvimento deste projeto de Business Intelligence permitiu aplicar, de forma prática, todo o ciclo de tratamento e análise de dados, desde a recolha, limpeza e modelação em SQL até à construção de dashboards analíticos em Power BI.

A nível analítico, os dados revelaram insights relevantes sobre os avistamentos de OVNI's. A **duração média dos avistamentos** foi de **21 minutos e 35 segundos**, e o número total de registos válidos ultrapassou os **74 mil**.

O mapa geográfico mostrou uma clara concentração de avistamentos nos **Estados Unidos**, sendo também os países com mais ocorrências o Canadá, o Reino Unido, a Austrália e a Alemanha.

No que respeita às formas mais avistadas, destacaram-se as categorias **teardrop**, **changed**, **fireball**, **delta** e **cone**. A análise temporal revelou um crescimento acentuado dos avistamentos a partir de **1994**, atingindo um pico entre **2008 e 2012**, seguido de uma queda notável até **2014**.

A nível de padrões comportamentais, observou-se que a maioria dos avistamentos está classificada como de **duração média**, e que os **sábados** e **domingos** são os dias com maior incidência de relatos, o que pode estar relacionado com maior tempo livre, observação mais atenta ou fenómenos sociais.

Para além dos resultados da análise, este projeto foi um exercício desafiante de adaptação e resolução de problemas reais, nomeadamente a instabilidade da ligação entre Power BI e MySQL, que obrigou à reformulação da abordagem técnica. A alternativa encontrada permitiu dar continuidade ao projeto e reforçou a importância da flexibilidade e resiliência em contexto de BI.





Concluimos que, mais do que procurar respostas definitivas, o verdadeiro valor da análise de dados está em **explorar padrões, levantar hipóteses e apoiar decisões** com base em evidência. E sim... talvez a verdade esteja mesmo lá fora.

Analizamos o desconhecido, enfrentamos desafios reais, e saímos mais fortes. Somos Area52 e estamos prontos para o próximo enigma!



## Bibliografia e Webgrafia

- [Dataset UFO sightings](#)
- [Manual DAX Microsoft](#)
- [Manual MySQL](#)



## Lista de anexos

[Anexo A - Ficheiro CSV - tabela mãe com primeira limpeza de dados](#)

[Anexo B - Script MySQL - criação, inserção e normalização de dados](#)

[Anexo C - Ficheiros CSV - extraídos da base de dados](#)

[Anexo D - Dashboard Power BI](#)