



Instituto Politécnico de Gestão e Tecnologia

Curso PG em *Analytics e Data Science Empresarial*

Análise da Eficiência Hospitalar no SNS: Aplicação de Business Intelligence à
Atividade de Internamento e Ocupação de Camas em Portugal (2013-2025)

Rita Noronha, 22503040

Joyce Silva Melo, 22502165

Docente:

Jorge Duque

Unidade Curricular: Fundamentos de Business Intelligence e Análise de Dados

Vila Nova de Gaia

Ano académico <2025-2026>

[página de verso, em branco]

Resumo

O presente trabalho analisa dados disponibilizados pelo Serviço Nacional de Saúde (SNS) com o objetivo de compreender a utilização dos internamentos hospitalares e a eficiência da ocupação de camas. Com base em dois datasets – “Atividade de Internamento Hospitalar” e “Ocupação do Internamento” – foram aplicadas técnicas de ETL em Power Query para garantir a integridade dos dados, tratar valores nulos e criar colunas auxiliares de qualidade. Posteriormente, desenvolveu-se um modelo dimensional em Power BI, seguindo práticas de modelação. As análises exploratórias permitiram observar indicadores como, entre outros, o número de internamentos por região, a média de dias por especialidade e a pressão hospitalar. Adicionalmente, aplicaram-se técnicas básicas de previsão (regressão linear) para estimar a evolução do número de internamentos e da lotação praticada. Conclui-se que o sistema hospitalar português apresenta uma operação globalmente consistente, embora com desafios pontuais. Este projeto constitui uma base sólida para análises futuras mais granulares, com integração de múltiplas fontes e aplicação de modelos preditivos mais robustos.

Palavras-chave: eficiência hospitalar; ETL; internamentos hospitalares; ocupação hospitalar; Power BI; previsão

INTRODUÇÃO

A crescente complexidade dos sistemas de saúde atuais — marcada pelo envelhecimento da população, pela variabilidade na procura de cuidados e pelos constrangimentos orçamentais — impõe desafios significativos à gestão eficiente de recursos hospitalares. Neste contexto, a adoção de abordagens baseadas em análise de dados (*data analytics*) e *Business Intelligence* (BI) revela-se estratégica para melhorar a alocação de recursos, aumentar a eficiência operacional e apoiar decisões informadas de saúde pública (Worku, 2025; Nwoke, 2024). A integração de dados sobre internamentos e ocupação de camas hospitalares permite não só avaliar o uso atual da capacidade instalada, como também antecipar necessidades futuras, contribuindo para a sustentabilidade e resiliência dos hospitais. Além disso, compreender a disponibilidade e a ocupação das camas hospitalares é fundamental para qualquer utilizador do sistema de saúde. A existência de camas suficientes significa acesso mais rápido ao atendimento, menor risco de atrasos em cirurgias ou internamentos urgentes e maior qualidade na prestação dos cuidados. Quando a taxa de ocupação é elevada, os serviços tornam-se mais lentos, a capacidade de resposta diminui e o risco de sobrecarga aumenta. Por isso, analisar esses dados oferece valor direto para o cidadão, permitindo entender como a organização dos recursos hospitalares influencia o atendimento que recebe. Note-se que estes dados podem apoiar múltiplas vertentes da gestão hospitalar, incluindo o planeamento da capacidade de internamento, a alocação de recursos humanos e técnicos, a identificação de áreas prioritárias para investimento, a monitorização de eficiência institucional e o suporte à tomada de decisão baseada em evidência. Este trabalho assenta na exploração de dois conjuntos de dados públicos disponibilizados pelo Serviço Nacional de Saúde (SNS): por um lado, a “Atividade de Internamento Hospitalar”; por outro, a “Ocupação do Internamento”. Através de um processo de ETL (*Extract, Transform, Load*) em *Power Query*, procedeu-se à limpeza, normalização e verificação de consistência dos dados — incluindo tratamento de dados em falta — garantindo a fiabilidade da base analítica. Posteriormente, foi construído um modelo dimensional no *Power BI*, a partir do qual foram gerados indicadores como média de dias de internamento, taxa de ocupação de camas e pressão hospitalar por região e especialidade. Para além da análise descritiva, recorreu-se a técnicas preditivas — nomeadamente regressão linear — com o objetivo de projetar tendências de internamentos e utilização de camas para os próximos anos, antecipando cenários de sobrecarga ou subutilização. Tal abordagem confirma o potencial dos métodos de *healthcare data analytics* para apoiar a tomada de decisão estratégica e o planeamento de capacidade (Nwoke, 2024; “*Healthcare data analytics and predictive*

modelling”, 2024). Em suma, este estudo pretende demonstrar que, com dados públicos e técnicas acessíveis, é viável gerar conhecimento relevante para a gestão hospitalar. A análise proposta contribui para a discussão sobre eficiência e sustentabilidade do sistema de saúde público em Portugal e constitui uma base para trabalhos futuros envolvendo modelos preditivos mais sofisticados e múltiplas fontes de dados.

1. SELEÇÃO, CARACTERIZAÇÃO E ENQUADRAMENTO DOS *DATASETS* UTILIZADOS

A monitorização contínua da atividade hospitalar e da capacidade instalada constitui um elemento fundamental para a gestão eficiente dos sistemas públicos de saúde. No contexto do Serviço Nacional de Saúde (SNS) a análise sistemática dos internamentos e da ocupação hospitalar permite antecipar constrangimentos, alocar recursos de forma mais eficiente e garantir a continuidade dos cuidados prestados à população (Guedes, 2013; Batko & Ślęzak, 2022). O presente trabalho, aplicado à saúde, recorre a fontes de dados de elevada fiabilidade e abrangência, disponibilizadas pelo Portal da Transparência do SNS, com o objetivo de construir indicadores de desempenho hospitalar, analisar padrões regionais e estimar a evolução da procura hospitalar nos próximos anos.

1.1. Fontes de Dados

Foram selecionados dois conjuntos de dados disponibilizados pelo Portal da Transparência do SNS:

- Atividade de Internamento Hospitalar:
<https://transparencia.sns.gov.pt/explore/dataset/atividade-de-internamento-hospitalar>
- Ocupação do Internamento:
<https://transparencia.sns.gov.pt/explore/dataset/ocupacao-do-internamento>

Ambos os *datasets* são constituídos por dados administrativos reais reportados pelas unidades hospitalares do SNS, e estão disponíveis para consulta pública, em formato estruturado.

1.2. Descrição e Estrutura dos *Datasets*

Dataset 1 – Atividade de Internamento Hospitalar

Este *dataset* descreve a atividade assistencial das instituições de saúde públicas, permitindo compreender a quantidade de episódios de internamento por ano, região, instituição e especialidade. Cada linha representa um registo agregado de uma unidade hospitalar num determinado ano e especialidade – Registos (linhas): 16.734; Variáveis (colunas): 7.

Tabela 1. Descrição das variáveis do *dataset* 1.

Variável	Descrição
Período	Ano do registo
Região	Região de Saúde (Norte, Centro, LVT, Alentejo, Algarve)
Instituição	Nome da unidade hospitalar
Localização Geográfica	Coordenadas geográficas da instituição
Tipo de Especialidade	Ex.: Cirúrgica, Médica
Doentes Saídos	Total de pacientes internados que receberam alta
Dias de Internamento	Total de dias de hospitalização correspondentes

Dataset 2 – Ocupação do Internamento

Este segundo *dataset* complementa a análise ao disponibilizar variáveis relativas à capacidade instalada e à sua utilização efetiva ao longo do ano. Cada linha representa a situação de uma instituição num dado ano – Registos (linhas): 7.126; Variáveis (colunas): 7.

Tabela 2. Descrição das variáveis do *dataset* 1.

Variável	Descrição
Período	Ano do registo
Região	Região de Saúde
Instituição	Nome da unidade hospitalar
Localização Geográfica	Coordenadas da instituição
Nº Dias de Internamento	Total de dias em que camas estiveram ocupadas
Lotação Praticada	Capacidade efetiva de camas disponíveis durante o ano
Taxa Anual de Ocupação	Percentagem de utilização das camas com base na capacidade instalada

1.3. Justificativa dos *Datasets*

A seleção destes dois *datasets* justifica-se por três razões principais. Primeiramente, destaca-se a disponibilidade pública e a credibilidade institucional das fontes, uma vez que os dados são disponibilizados pelo próprio Serviço Nacional de Saúde (SNS), com cobertura nacional e atualizações regulares. Em segundo lugar, salienta-se a amplitude e granularidade da informação fornecida, uma vez que os *datasets* abrangem vários anos, diversas especialidades médicas, o que possibilita a realização de análises temporais, geográficas e

clínicas robustas. Por fim, evidencia-se o seu potencial analítico e estratégico, na medida em que, quando analisados em conjunto, os dois conjuntos de dados permitem a construção de indicadores relevantes de eficiência fundamentais para o planeamento hospitalar, a previsão da procura e a identificação de pressões assistenciais nas instituições de saúde.

1.4. Complementaridade entre os *Datasets*

Os dois conjuntos de dados, quando analisados em conjunto, proporcionam uma visão integrada do funcionamento hospitalar. O primeiro incide sobre a procura hospitalar, traduzida pelo número de internamentos e a sua distribuição por região, especialidade e instituição. O segundo permite avaliar a capacidade de resposta das unidades hospitalares, através de indicadores como a lotação praticada, a ocupação real e a pressão exercida sobre os recursos disponíveis.

A análise combinada destes dados possibilita a identificação de regiões com situações de sobrecarga ou, pelo contrário, de subutilização de recursos, permitindo também comparar a eficiência relativa entre diferentes instituições hospitalares. Para além disso, esta abordagem integrada facilita a estimativa de tendências e a projeção de cenários futuros de ocupação, contribuindo de forma relevante para o planeamento estratégico da capacidade instalada e para a tomada de decisões informadas no âmbito da gestão hospitalar.

1.5. Qualidade dos Dados

Durante o processo de exploração inicial, foram identificados valores em falta (nulos) em variáveis críticas como “Doentes Saídos” e “Dias de Internamento”, o que exigiu uma abordagem cuidadosa na fase de pré-processamento. Foram criadas colunas auxiliares para assinalar registos com dados incompletos, sem substituir os nulos por valores arbitrários, garantindo a integridade analítica. Esta decisão metodológica será detalhada, seguidamente, na secção de ETL.

2. PROCESSO ETL (EXTRACT, TRANSFORM, LOAD)

O processo ETL constituiu a etapa fundamental para assegurar a qualidade, consistência e fiabilidade dos dados utilizados na construção do modelo. O procedimento envolveu quatro fases principais: (i) verificação e normalização dos tipos de dados; (ii) identificação e tratamento de valores nulos; (iii) padronização de texto e limpeza semântica; e (iv) criação de colunas derivadas para controlo da qualidade da informação. Todo o processamento foi realizado em *Power Query*, garantindo a rastreabilidade das transformações e a preservação da integridade das bases originais.

2.1. Verificação e Correção dos Tipos de Dados

A primeira etapa consistiu em validar se o *Power Query* interpretou corretamente os tipos de dados de cada coluna. Este passo é essencial, uma vez que tipos incorretos podem distorcer cálculos, impedir a criação de relações corretas no modelo dimensional ou provocar erros em funções DAX.

Após a importação dos *datasets*, foi verificado que ambos já apresentavam os tipos de dados corretos, conforme descrito:

Dataset 1 – Atividade de Internamento Hospitalar:

- Período – Data;
- Região – Texto;
- Instituição – Texto;
- Localização Geográfica – Texto;
- Tipo de Especialidade – Texto;
- Doentes Saídos – Número inteiro;
- Dias de Internamento – Número inteiro.

Dataset 2 – Ocupação do Internamento:

- Período – Data;
- Região – Texto;
- Instituição – Texto;
- Localização Geográfica – Texto;
- Nº Dias de Internamento – Número inteiro;
- Lotação Praticada – Número inteiro;
- Taxa Anual de Ocupação – Número decimal.

2.2. Verificação da Existência de Linhas Vazias

Para avaliar a integridade da informação, foram ativadas em “View” as ferramentas *Column Quality*, *Column Distribution* e *Column Profile*, que permitem visualizar graficamente a quantidade de valores válidos, nulos e erros.

Resultados obtidos:

Dataset 1 – Atividade de Internamento Hospitalar:

- 0% de valores vazios em colunas de texto e data;
- Menos de 1% de valores nulos em colunas numéricas.

Como uma linha vazia exigiria que todas as colunas tivessem valor “Empty”, concluiu-se que o *dataset* não possui linhas completamente vazias, apenas alguns valores nulos isolados, que serão tratados na etapa seguinte.

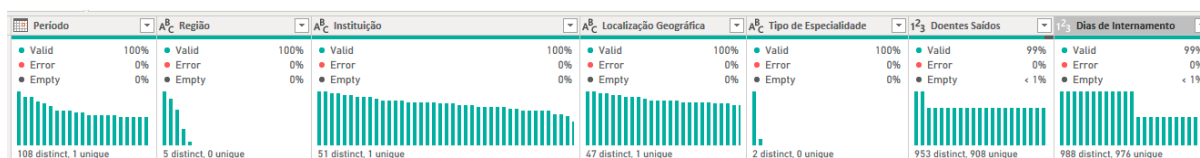


Figura 1. Verificação da existência de linhas vazias – Dataset 1.

Dataset 2 – Ocupação do Internamento:

- 0% de valores vazios nas colunas de texto e data;
- Menos de 1% de valores nulos nas colunas *Nº Dias de Internamento*, *Lotação Praticada* e *Taxa de Ocupação*.

Como uma linha vazia exigiria que todas as colunas estivessem marcadas como “Empty”, foi possível inferir que o *dataset* não possui linhas completamente vazias, apenas alguns valores nulos isolados que serão tratados na etapa seguinte.

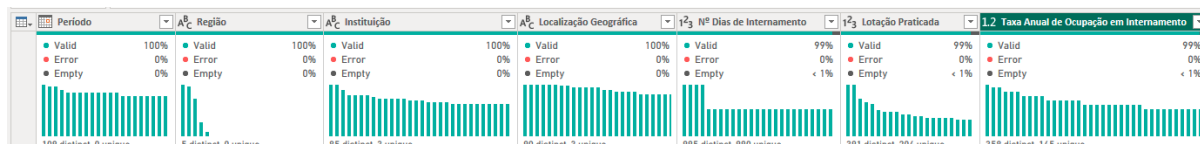


Figura 2. Verificação da existência de linhas vazias – Dataset 2.

2.3. Padronização de Textos com *Trim* e *Clean*

Para garantir consistência semântica e evitar falhas nos *joins* do modelo dimensional, aplicaram-se as funções *Trim* e *Clean* a todas as colunas de texto nos dois *datasets*.

- *Trim* remove espaços no início e fim das *strings*;
- *Clean* elimina caracteres invisíveis que comprometem agrupamentos ou filtros.

Deste modo, no *dataset* 1 e 2, foram utilizadas as funções *Trim* e *Clean* do *Power Query* para padronizar todas as colunas de texto. Essas funções removem espaços desnecessários e caracteres invisíveis, garantindo consistência entre os dados e evitando problemas nos relacionamentos, filtros e gráficos do modelo.

Este procedimento é crítico para assegurar que duas instituições com nomes aparentemente iguais não sejam tratadas como entidades distintas devido a caracteres escondidos.

3. TRATAMENTO DE VALOR NULOS

Dataset 1 – Atividade de Internamento Hospitalar:

Durante o processo de ETL, identificámos a presença de valores nulos nas colunas Doentes Saídos e Dias de Internamento. Antes de decidir substituí-los por zero, procedeu-se à análise para entender o que esses *nulls* realmente representavam no contexto operacional do *dataset*.

No caso destas duas métricas, o valor 0 possui um significado real e específico:

- Doentes Saídos - Indica que nenhum doente teve alta naquele período;
- Dias de Internamento - Indica que não houve qualquer dia contabilizado.

Por outro lado, o valor *null* representa ausência de informação, e não ausência de atividade. Por isso, substituir *null* por zero sem avaliar o contexto poderia gerar interpretações incorretas, como sugerir que instituições não tiveram movimento quando, na verdade, apenas não registaram os dados.

Para fundamentar a decisão, procedeu-se à seguinte análise:

- Quantidade e distribuição dos valores nulos;
- Possibilidade de se concentrarem em determinadas instituições ou períodos;
- Se apareciam acompanhados de outros campos vazios na mesma linha;
- Se faria sentido existir realmente “zero” para aquele registo.

A análise mostrou que os *nulls* existentes eram reduzidos e surgiam de forma isolada, sem indicar ausência real de doentes ou de dias de internamento. Assim, concluiu-se que se tratava de registos incompletos e não valores válidos.

Com base nisso, decidiu-se não substituir *null* por zero no ETL, preservando a integridade da base original e evitando distorções nos totais e médias. O tratamento será feito apenas na fase de DAX, quando necessário, utilizando funções que substituem *null* apenas na apresentação dos resultados.

Para apoiar esta avaliação, criámos colunas auxiliares que identificam automaticamente quando cada uma dessas métricas está em falta (“Dados Em Falta Saídos”, “Dados Em Falta Internamento”), permitindo monitorizar facilmente os registos incompletos ao longo da análise.

```
if [Doentes Saídos] = null then "Sim" else "Não"

if [Dias de Internamento] = null then "Sim" else "Não"
```

Custom Column

Add a column that is computed from the other columns.

New column name

Dados Em Falta Saídos

Custom column formula ⓘ

= if [Doentes Saídos] = null then "Sim" else "Não"

Available columns

região
Instituição
Localização Geográfica
Tipo de Especialidade
Doentes Saídos
Dias de Internamento
Dados Em Falta Saídos
Dados Em Falta Internamento

Figura 3. Função da coluna auxiliar “Dados Em Falta Saídos”.

[illegible]

Figura 4. Visualização da coluna auxiliar “Dados Em Falta Saídos”.

À semelhança do primeiro *dataset*, verificámos que os valores nulos presentes nas colunas numéricas (*Nº Dias de Internamento*, *Lotação Praticada* e *Taxa Anual de Ocupação*) não representam valores reais iguais a zero, mas apenas dados não registados.

Como os *nulls* são poucos e surgem de forma isolada, inferiu-se que se tratam de registos incompletos, e não de ausência de atividade. Por isso, optou-se por não substituir *null* por zero no ETL, mantendo a base original íntegra.

Procedeu-se, assim, igualmente, à criação de colunas auxiliares para identificar facilmente estes casos:

- *DadosEmFalta_DiasInternamento*
- *DadosEmFalta_LotacaoPraticada*
- *DadosEmFalta_TaxaOcupacao*

4. ARQUITETURA BI E MODELO DE DADOS

4.1. Visão geral da Arquitetura de *Business Intelligence*

A arquitetura de *Business Intelligence* concebida para este projeto segue uma estrutura simplificada, dividida em três camadas fundamentais:

- A. Camada de Fontes de Dados: constituída pelos *datasets* extraídos do Portal da Transparência do Serviço Nacional de Saúde, nomeadamente: *Atividade de Internamento Hospitalar e Ocupação do Internamento*;
- B. Camada de *Staging*/Preparação: na qual foram aplicadas todas as transformações no *Power Query*, conforme descrito no capítulo anterior. Esta camada inclui:
 - Validação de tipos de dados;
 - Tratamento de valores nulos;
 - Criação de colunas auxiliares;
 - Normalização textual.
- C. Camada Analítica: representada pelo modelo dimensional implementado no *Power BI*. Nesta fase, as tabelas são organizadas num modelo em estrela (*star schema*), permitindo análises rápidas, flexíveis e intuitivas por múltiplas perspetivas (tempo, região, instituição, especialidade e localização).

4.2. Construção do Modelo Dimensional

A camada analítica do projeto foi construída com um modelo dimensional, no formato *star schema*, para tornar a análise dos dados mais organizada e eficiente. Nesse modelo, as tabelas de factos que incluem informações numéricas como número de internamentos, dias de internamento e ocupação hospitalar ficam no centro. Circunscritas encontram-se as tabelas de dimensões, como *Calendário*, *Região*, *Instituição*, *Especialidade* e *Localização*, que fornecem os detalhes necessários para filtrar e explicar esses números. As relações seguem o padrão “um para muitos”, garantindo que os filtros funcionem corretamente no relatório. Esta estrutura facilita a exploração dos dados, permitindo comparar anos, regiões, especialidades e instituições de forma simples e rápida.

4.2.1. Tabela *Dim_Calendario*

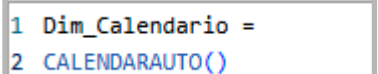
Após a conclusão do processo de ETL, iniciámos a fase de modelagem dimensional do projeto. O primeiro passo consistiu na criação da Dimensão Calendário (*Dim_Calendario*), fundamental para garantir análises temporais consistentes entre os dois *datasets* utilizados:

- Atividade de Internamento Hospitalar
- Ocupação do Internamento

A *Dim_Calendario* foi criada com DAX, gerando uma tabela de datas contínuas com base no intervalo temporal presente nas tabelas de factos. Foram adicionadas colunas derivadas para suportar análises por diferentes níveis temporais:

- Ano
- Mês (número)
- Mês (nome)
- Ano-Mês
- Trimestre

Esta tabela foi marcada como tabela de data no *Power BI*, assegurando que todas as funções temporais operam corretamente.



```
1 Dim_Calendario =  
2 CALENDAR(AUTO())
```

Figura 6. Criação da Tabela *Dim_Calendario*.

4.2.2. Tabela *Dim_Regiao*

A *Dim_Regiao* foi construída para centralizar a informação relativa às regiões de saúde, que se encontram repetidas nos dois *datasets*. Esta dimensão garante que os filtros por região afetam simultaneamente as duas tabelas de factos, permitindo comparações consistentes entre internamentos e ocupação hospitalar.

```
1 Dim_Regiao =
2 DISTINCT (
3     UNION (
4         SELECTCOLUMNS (
5             'atividade-de-internamento-hospitalar',
6             "Regiao", 'atividade-de-internamento-hospitalar'[Região]
7         ),
8         SELECTCOLUMNS (
9             'ocupacao-do-internamento',
10            "Regiao", 'ocupacao-do-internamento'[Região]
11        )
12    )
13 )
```

Figura 7. Criação da Tabela *Dim_Regiao*.

4.2.3. Tabela *Dim_Instituicao*

A dimensão *Dim_Instituicao* agrega todas as instituições hospitalares (hospitais, ULS, centros hospitalares) envolvidas nos registos. Esta dimensão permite aplicar filtros únicos por hospital, evitando duplicações e assegurando coerência na análise cruzada dos dois *datasets*.

```
1 Dim_Instituicao =
2 DISTINCT (
3     UNION (
4         SELECTCOLUMNS (
5             'atividade-de-internamento-hospitalar',
6             "Instituicao", 'atividade-de-internamento-hospitalar'[Instituição]
7         ),
8         SELECTCOLUMNS (
9             'ocupacao-do-internamento',
10            "Instituicao", 'ocupacao-do-internamento'[Instituição]
11        )
12    )
13 )
```

Figura 8. Criação da Tabela *Dim_Instituicao*.

4.2.4. Tabela Dim_Especialidade

A *Dim_Especialidade* foi extraída exclusivamente da tabela *Fato_Internamentos*, pois apenas esta contém a variável “*Tipo de Especialidade*”. Foram consideradas categorias como *Cirúrgica* e *Médica*. Esta dimensão facilita a análise segmentada da atividade hospitalar por tipologia clínica.

```
1 Dim_Especialidade =  
2 DISTINCT (  
3 |   'atividade-de-internamento-hospitalar'[Tipo de Especialidade]  
4 )  
-
```

Figura 9. Criação da Tabela *Dim_Especialidade*.

4.2.5. Tabela Dim_Localizacao Geográfica

A dimensão *Dim_Localizacao* reúne os dados relativos à localização geográfica (concelho ou município) das instituições. Esta centralização permite criar mapas e realizar análises regionais detalhadas, garantindo coerência nos filtros aplicados a ambas as tabelas de factos.

```
1 Dim_Localizacao =  
2 DISTINCT (  
3 |   UNION (  
4 |     SELECTCOLUMNS (  
5 |       'atividade-de-internamento-hospitalar',  
6 |       "Localizacao", 'atividade-de-internamento-hospitalar'[Localização Geográfica]  
7 |     ),  
8 |     SELECTCOLUMNS (  
9 |       'ocupacao-do-internamento',  
10 |      "Localizacao", 'ocupacao-do-internamento'[Localização Geográfica]  
11 |     )  
12 |   )  
13 )  
-
```

Figura 10. Criação da Tabela *Dim_Localizacao*.

4.3. Relacionamentos entre Tabelas Fato

Todos os relacionamentos do modelo seguem a estrutura “*One-to-Many*”, com direção de filtro Single a partir das tabelas de dimensão para as tabelas de factos.

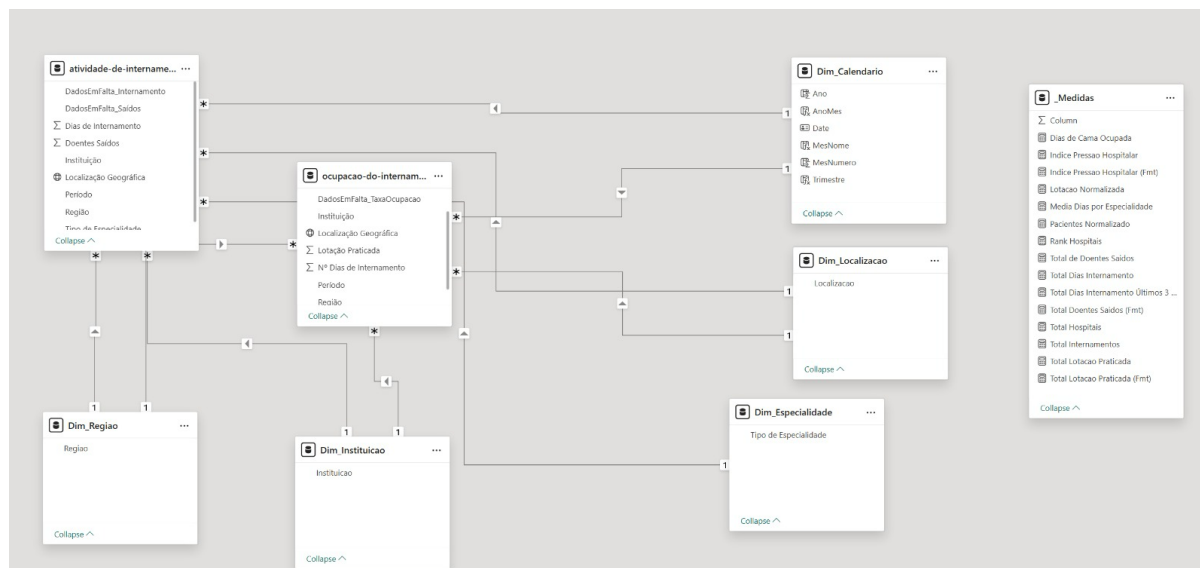


Figura 11. Modelo Dimensional.

Esta configuração assegura que os filtros aplicados em *dashboards* e visualizações são corretamente propagados, sem gerar ambiguidade.

Tabela 2. Descrição das variáveis do *dataset* 1.

Tabela de Dimensão	Relacionada com...	Chave de Ligação
<i>Dim_Calendarario</i>	<i>Fato_Internamentos</i> e <i>Fato_Ocupacao</i>	Período (Ano)
<i>Dim_Regiao</i>	Ambas as tabelas de factos	Região
<i>Dim_Instituicao</i>	Ambas as tabelas de factos	Instituição
<i>Dim_Especialidade</i>	Apenas <i>Fato_Internamentos</i>	Tipo de Especialidade
<i>Dim_Localizacao</i>	Ambas as tabelas de factos	Localização Geográfica

Manage relationships



+ New relationship		⚡ Autodetect			✎ Edit	🗑 Delete	≡ Filter	▼
<input type="checkbox"/>	From: table (column)	↑	Relationship	To: table (column)	Status			
<input type="checkbox"/>	atividade-de-internamento-ho...		* — ◀ — 1	Dim_Instituicao (Instituicao)	Active	...		
<input type="checkbox"/>	atividade-de-internamento-ho...		* — ◀ — 1	Dim_Localizacao (Localizacao)	Active	...		
<input type="checkbox"/>	atividade-de-internamento-ho...		* — ◀ — 1	Dim_Calendarario (Date)	Active	...		
<input type="checkbox"/>	atividade-de-internamento-ho...		* — ◀ — 1	Dim_Regiao (Regiao)	Active	...		
<input type="checkbox"/>	atividade-de-internamento-ho...		* — ◀ — 1	Dim_Especialidade (Tipo de Es...	Active	...		
<input type="checkbox"/>	ocupacao-do-internamento (I...		* — ◀ — 1	Dim_Instituicao (Instituicao)	Active	...		
<input type="checkbox"/>	ocupacao-do-internamento (L...		* — ◀ — 1	Dim_Localizacao (Localizacao)	Active	...		
<input type="checkbox"/>	ocupacao-do-internamento (P...		* — ◀ — 1	Dim_Calendarario (Date)	Active	...		
<input type="checkbox"/>	ocupacao-do-internamento (R...		* — ◀ — 1	Dim_Regiao (Regiao)	Active	...		

Figura 12. Demonstração das relações estabelecidas no modelo.

4.4. Criação de Medidas com DAX

Para potenciar a análise de desempenho hospitalar, foram criadas medidas DAX relevantes, como:

Média de Dias por Doente =

*DIVIDE(SUM(Fato_Internamentos[Dias de Internamento]),
SUM(Fato_Internamentos[Doentes Saídos]))*

Taxa de Ocupação (%) =

*DIVIDE(SUM(Fato_Ocupacao[Nº Dias de Internamento]),
SUM(Fato_Ocupacao[Lotação Praticada]))*

Adicionalmente, foram criadas medidas complementares para análise por ano, região e especialidade, incluindo projeções com base em regressão linear (detalhadas em capítulo próprio).

5. *ANALYTICS E DATA MINING*

Análise Preditiva: Previsão de Número de Internamentos para 2026

Nesta análise foi utilizado o *dataset* de internamentos hospitalares para estudar a evolução do número de “Doentes Saídos” ao longo dos anos e estimar o valor previsto para 2026. A coluna analisada representa o total anual de internamentos concluídos em Portugal, agregando toda a atividade hospitalar do Serviço Nacional de Saúde. Ao observar a série histórica, verificaram-se algumas oscilações naturais, bem como a presença de valores nulos e uma redução acentuada em 2025, que provavelmente reflete dados ainda incompletos para esse ano.

Para preparar a informação, a coluna “Período” foi convertida para o formato de data e os valores foram agregados por ano através do somatório de internamentos. Esta transformação permitiu obter uma visão anual mais clara e facilitar a identificação da tendência geral. Com base nessa série temporal, aplicou-se um modelo simples de Regressão Linear, adequado a análises introdutórias e suficiente para estimar a evolução dos internamentos de forma aproximada.

Após o treino do modelo com os valores históricos, foi gerada uma estimativa para o ano seguinte. O valor previsto foi de aproximadamente 3,79 milhões de internamentos para 2026, refletindo uma tendência de continuidade relativamente aos anos anteriores, embora influenciada pela quebra registada em 2025. O gráfico criado destaca os valores reais e o ponto previsto, permitindo visualizar de forma clara a evolução da atividade hospitalar ao longo do tempo.

Embora simples, esta previsão oferece um indicador útil para compreender a direção geral da procura por cuidados de internamento e pode servir como ponto de partida para análises mais detalhadas. Trata-se de uma abordagem exploratória, cujo objetivo é demonstrar a aplicação de técnicas básicas de *Data Mining* no contexto da saúde, complementando a análise descritiva realizada no *Power BI*.

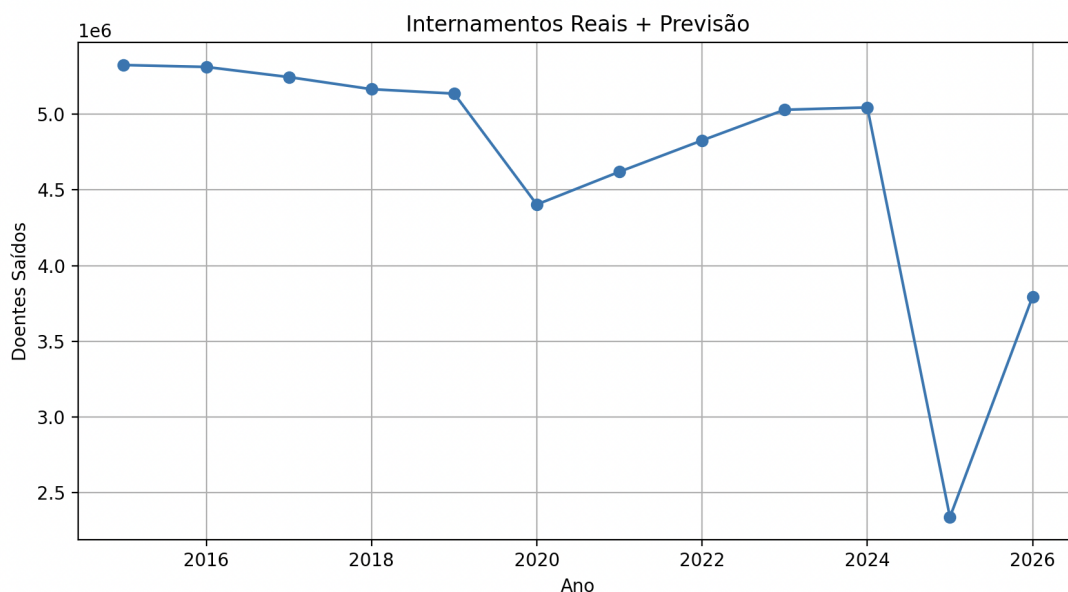


Figura 13. Previsão do N.º de Internamentos com Regressão Linear

Conforme referido, o gráfico apresentado foi elaborado com recurso a *Python*, representando a evolução histórica do número de doentes saídos entre 2015 e 2025, com a estimativa para o ano de 2026 calculada através de um modelo de regressão linear simples. A previsão para 2026 foi de 3.794.926 internamentos, tendo em conta a tendência dos anos anteriores, incluindo o impacto da pandemia de COVID-19 e a subsequente recuperação.

```
~/trabalho python BI on v28.4.0 (desktop-linux) via trabalho python BI via system
+ "/Users/joyce/trabalho python BI/.venv/bin/python" "/Users/joyce/trabalho python BI/import pandas as pd.py"

Internamentos por Ano:
  Ano  Doentes Saídos
0  2015    5325372.0
1  2016    5312616.0
2  2017    5245699.0
3  2018    5165716.0
4  2019    5136546.0
5  2020    4403864.0
6  2021    4620206.0
7  2022    4827198.0
8  2023    5029895.0
9  2024    5045081.0
10 2025    2338668.0
/Users/joyce/trabalho python BI/.venv/lib/python3.13/site-packages/sklearn/utils/validation.py:2749: UserWarning: X does
not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(

Previsão para o ano 2026: 3794926
```

Figura 14. Código e Saída da Regressão Linear em *Python*.

A figura representa um excerto do ambiente de execução do *script Python*, desta forma, evidencia-se a utilização da biblioteca *sklearn* para aplicar regressão linear sobre os dados anuais de internamentos. A previsão foi gerada a partir dos valores históricos, com o objetivo de apoiar análises prospectivas no contexto da gestão hospitalar.

```

previsao_ocupacao.py
1 import pandas as pd
2 from sklearn.linear_model import LinearRegression
3 import matplotlib.pyplot as plt
4
5 # 1. Carregar o dataset
6 df = pd.read_csv(
7     "/Users/joyce/Downloads/ocupacao-do-internamento.csv",
8     sep=";"
9 )
10
11 # 2. Converter Período para data
12 df["Período"] = pd.to_datetime(df["Período"], format="%Y-%m", errors="coerce")
13
14 # 3. Criar coluna Ano
15 df["Ano"] = df["Período"].dt.year
16
17 # 4. Agregar a Lotação Praticada por ano
18 col_ocupacao = "Lotação Praticada" # altere se o nome for diferente
19 df_ano = df.groupby("Ano")[col_ocupacao].mean().reset_index()
20
21 print("\nLotação média por ano:")
22 print(df_ano)
23
24 # 5. Preparar modelo
25 X = df_ano[["Ano"]]
26 y = df_ano[col_ocupacao]
27
28 modelo = LinearRegression()
29 modelo.fit(X, y)
30
31 # 6. Prever o próximo ano
32 proximo_ano = int(df_ano["Ano"].max() + 1)
33 previsao = float(modelo.predict([[proximo_ano]])[0])
34
35 print(f"\nPrevisão da Lotação Praticada para {proximo_ano}: {previsao:.2f}")
36
37 # 7. Criar dataframe final
38 df_prev = df_ano.copy()
39 df_prev.loc[len(df_prev)] = [proximo_ano, previsao]
40
41 # 8. Salvar previsão
42 df_prev.to_csv("/Users/joyce/Downloads/previsao_ocupacao.csv", index=False)
43 print("\nArquivo 'previsao_ocupacao.csv' criado!")
44
45 # 9. Gráfico com destaque
46 plt.figure(figsize=(12,6))
47
48 plt.plot(df_prev["Ano"][:-1], df_prev[col_ocupacao][:-1],
49         marker="o", label="Valores Reais")
50
51 plt.plot(df_prev["Ano"].iloc[-1], df_prev[col_ocupacao].iloc[-1],
52         marker="o", color="red", markersize=10, label=f"Previsão {proximo_ano}")
53
54 plt.text(
55     df_prev["Ano"].iloc[-1],
56     df_prev[col_ocupacao].iloc[-1] + df_prev[col_ocupacao].mean() * 0.03,
57     f"{df_prev[col_ocupacao].iloc[-1]:.2f}",
58     ha="center", fontsize=12, color="red"
59 )
60
61 plt.title("Lotação Praticada - Valores Reais + Previsão")
62 plt.xlabel("Ano")
63 plt.ylabel("Lotação Praticada (Média Anual)")
64 plt.grid(True)
65 plt.legend()
66 plt.tight_layout()
67 plt.show()
68
69 # 10. Mostrar tabela final
70 print("\nTabela final com previsão:")
71 print(df_prev)
72

```

Figura 15. Código *Python* para Previsão da Lotação Praticada com Regressão Linear.

A figura ilustra o *script* desenvolvido em *Python* para calcular a previsão da lotação praticada média anual com base em dados históricos do *dataset Ocupação do Internamento*. O código recorre à biblioteca *scikit-learn* para aplicar um modelo de regressão linear simples, bem como à biblioteca *matplotlib* para gerar o gráfico correspondente. A previsão é calculada para o ano seguinte ao último disponível no ficheiro CSV original, sendo ainda incluída a exportação da tabela com os resultados.

Análise Preditiva: Previsão de Ocupação para 2026

Nesta análise foi utilizado o *dataset* de ocupação do internamento para estudar a evolução da “*Lotação Praticada*” e estimar o valor previsto para o ano seguinte. A “*Lotação Praticada*” corresponde ao número médio diário de camas ocupadas nos serviços de internamento dos hospitais do Serviço Nacional de Saúde. Assim, cada valor representa quantos doentes se encontram internados, em média, por dia, considerando toda a rede hospitalar pública de Portugal.

Para preparar os dados, a coluna “*Período*” foi convertida para o formato de data e os valores foram agregados por ano através da média anual. Esta transformação permitiu obter uma visão mais estável da ocupação hospitalar, reduzindo a variabilidade mensal e facilitando a identificação da tendência geral ao longo dos anos.

Com esta série temporal anual, aplicou-se um modelo simples de Regressão Linear, adequado a análises introdutórias e suficiente para estimar tendências de evolução. O modelo foi treinado com os valores históricos e, a partir disso, gerou uma estimativa para o ano seguinte. O valor previsto foi de 466.46 camas ocupadas por dia, sugerindo uma tendência de estabilidade na lotação praticada, sem aumentos bruscos face aos anos anteriores.

O gráfico produzido destaca os valores reais e o ponto previsto, permitindo visualizar de forma clara a evolução da ocupação hospitalar. Embora a previsão seja baseada num método simples, fornece um indicador útil para a gestão, ajudando a antecipar necessidades de capacidade e apoiar decisões relacionadas com planeamento de recursos. Trata-se de uma análise exploratória, mas suficientemente robusta para demonstrar a aplicação de técnicas básicas de *Data Mining* em contexto hospitalar.

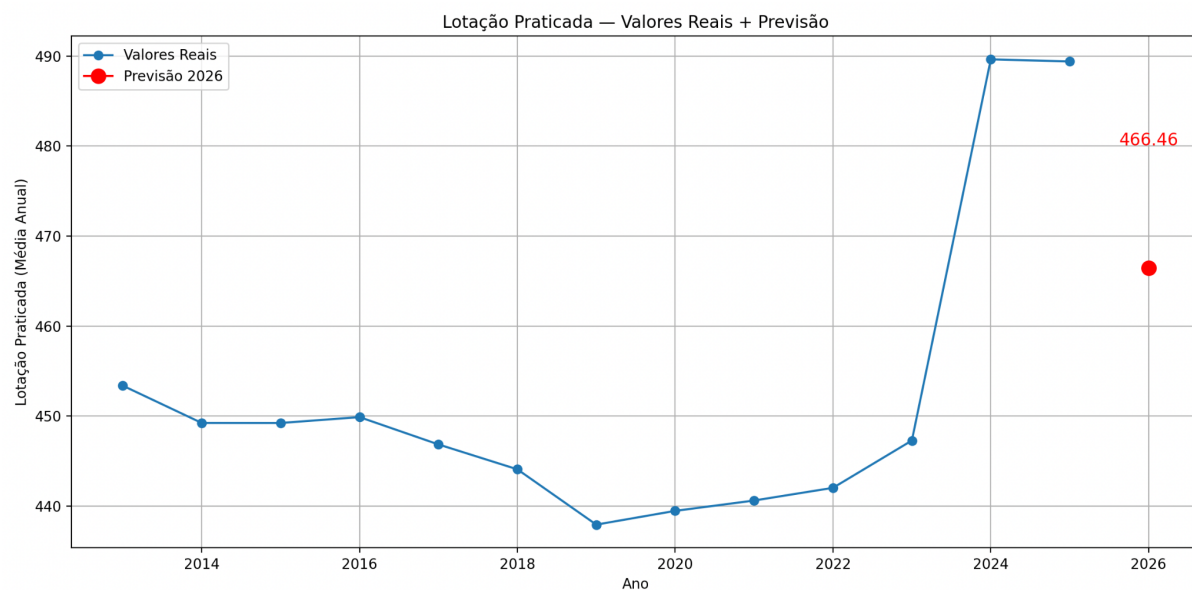


Figura 16. Previsão da Lotação Praticada para 2026 com Regressão Linear.

Gráfico representativo da evolução da lotação praticada média anual nas instituições hospitalares do SNS entre 2013 e 2025, com projeção para o ano de 2026 (vermelho), estimada em 466,46 unidades. O gráfico permite visualizar tendências ao longo do tempo, incluindo a quebra nos anos de pandemia e a recuperação recente.

```
Lotação média por ano:
  Ano  Lotação Praticada
0  2013      453.385027
1  2014      449.244165
2  2015      449.239362
3  2016      449.893428
4  2017      446.858639
5  2018      444.093750
6  2019      437.949653
7  2020      439.477431
8  2021      440.626736
9  2022      442.036713
10 2023      447.292419
11 2024      489.635478
12 2025      489.396501
/Users/joyce/trabalho python BI/.venv/lib/python3
.13/site-packages/sklearn/utils/validation.py:274
9: UserWarning: X does not have valid feature nam
es, but LinearRegression was fitted with feature
names
  warnings.warn(
Previsão da Lotação Praticada para 2026: 466.46
```

Figura 17. Saída da Previsão da Lotação Praticada em *Python*.

A figura representa o terminal com os valores médios anuais de lotação praticada calculados para o período de 2013 a 2025, com destaque para a previsão de 466,46 em 2026. A execução foi realizada com *Python*.

```
previsao_ocupacao.py > ...
1  import pandas as pd
2  from sklearn.linear_model import LinearRegression
3  import matplotlib.pyplot as plt
4
5  # 1. Carregar o dataset
6  df = pd.read_csv(
7      "/Users/joyce/Downloads/ocupacao-do-internamento.csv",
8      sep=";"
9  )
10
11 # 2. Converter Período para data
12 df["Período"] = pd.to_datetime(df["Período"], format="%Y-%m", errors="coerce")
13
14 # 3. Criar coluna Ano
15 df["Ano"] = df["Período"].dt.year  Cannot access attribute "year" for class "Properties" Attribute "year"
16
17 # 4. Agregar a Lotação Praticada por ano
18 col_ocupacao = "Lotação Praticada" # altere se o nome for diferente
19 df_ano = df.groupby("Ano")[col_ocupacao].mean().reset_index()
20
21 print("\nLotação média por ano:")
22 print(df_ano)
23
24 # 5. Preparar modelo
25 X = df_ano[["Ano"]]
26 y = df_ano[col_ocupacao]
27
28 modelo = LinearRegression()
29 modelo.fit(X, y)
30
31 # 6. Prever o próximo ano
32 proximo_ano = int(df_ano["Ano"].max() + 1)
33 previsao = float(modelo.predict([proximo_ano])[0])  Argument of type "list[list[int]]" cannot be assigne
34
35 print(f"\nPrevisão da Lotação Praticada para {proximo_ano}: {previsao:.2f}")
36
37 # 7. Criar dataframe final
38 df_prev = df_ano.copy()
39 df_prev.loc[len(df_prev)] = [proximo_ano, previsao]
40
41 # 8. Salvar previsão
42 df_prev.to_csv("/Users/joyce/Downloads/previsao_ocupacao.csv", index=False)
43 print("\nArquivo 'previsao_ocupacao.csv' criado!")
44
45 # 9. Gráfico com destaque
46 plt.figure(figsize=(12,6))
47
48 plt.plot(df_prev["Ano"][:-1], df_prev[col_ocupacao][:-1],
49          markers="o", label="Valores Reais")
50
51 plt.plot(df_prev["Ano"].iloc[-1], df_prev[col_ocupacao].iloc[-1],
52          markers="o", color="red", markersize=10, label=f"Previsão (proximo_ano)")
53
54 plt.text(
55     df_prev["Ano"].iloc[-1],
56     df_prev[col_ocupacao].iloc[-1] + df_prev[col_ocupacao].mean() * 0.03,
57     f"{df_prev[col_ocupacao].iloc[-1]:.2f}",
58     ha="center", fontsize=12, color="red"
59 )
60
61 plt.title("Lotação Praticada - Valores Reais + Previsão")
62 plt.xlabel("Ano")
63 plt.ylabel("Lotação Praticada (Média Anual)")
64 plt.grid(True)
65 plt.legend()
66 plt.tight_layout()
67 plt.show()
68
69 # 10. Mostrar tabela final
70 print("\nTabela final com previsão:")
71 print(df_prev)
72
```

Figura 18. Script em *Python* para Previsão da Lotação Praticada.

Código *Python* desenvolvido para estimar a lotação praticada média anual no âmbito da análise hospitalar, utilizando a biblioteca *scikit-learn* para aplicar um modelo de regressão linear. O *script* inclui as etapas de carregamento do *dataset*, transformação da data, agregação anual, criação do modelo, previsão para o ano seguinte e visualização gráfica dos resultados.

6. DASHBOARDS EM POWER BI

Por forma a melhor enquadrar o trabalho em apreço, estruturámos a análise a partir de perguntas-chave que refletem as necessidades de profissionais de saúde e gestores ao interpretar dados sobre internamentos e capacidade hospitalar.

A exploração dos dados será estruturada e visualizada através de um *dashboard* interativo construído no *Power BI*.

Questões de Negócio:

1. Quantos episódios de internamento ocorreram por ano em Portugal?
2. Qual é o tempo médio de internamento por especialidade e quais áreas apresentam maior permanência?
3. Como estão distribuídos os internamentos pelas diferentes especialidades hospitalares?
4. Quais regiões apresentam maior pressão hospitalar, considerando simultaneamente o volume de pacientes internados e a utilização das camas?
5. Qual é a evolução da carga assistencial ao longo dos anos (internamentos \times dias de cama ocupada)?
6. Qual a participação de cada região na ocupação total de camas entre 2013 e 2025?
7. Qual especialidade é responsável pela maior carga assistencial (total de dias de internamento) nos últimos três anos?
8. Qual será a evolução do número de doentes internados nos próximos períodos?
9. Qual será a evolução da Lotação praticada nos próximos períodos?

Nota sobre valores nulos:

Ambos os *datasets* apresentam alguns valores nulos em variáveis relevantes. Para garantir transparência no processo analítico, foram criadas colunas auxiliares que identificam esses registos. Contudo, essas colunas não foram utilizadas nas visualizações nem nos modelos, uma vez que o número de ocorrências é reduzido e não altera de forma significativa as tendências globais analisadas.

A decisão de manter os valores em falta sem imputação deve-se ao carácter exploratório do trabalho. Em estudos futuros, recomenda-se avaliar o impacto destes registos incompletos e considerar técnicas de tratamento de dados para reforçar a precisão das análises.

Quantos episódios de internamento ocorreram por ano em Portugal?

O gráfico apresenta a evolução anual do número total de Doentes Saídos, indicador que representa o número de episódios de alta registados no sistema hospitalar português. A análise evidencia um volume consistente de internamentos entre 2015 e 2019, sempre em torno dos 5 milhões de episódios anuais.

Em 2020 observa-se uma redução significativa, alinhada com o impacto da pandemia na atividade hospitalar, que levou à diminuição de internamentos programados. Nos anos seguintes verifica-se uma recuperação gradual, aproximando-se novamente dos níveis pré-pandemia. O ano de 2025 apresenta um valor inferior por se tratar de dados ainda incompletos no *dataset*.

A soma total dos episódios de alta analisados ultrapassa 52 milhões, o que confirma a robustez e a qualidade da informação tratada, garantindo confiança nas análises subsequentes.

Considere-se que os 52 milhões se referem ao total de episódios de internamento registados, e não ao número de pacientes. Um mesmo doente pode gerar vários episódios ao longo dos anos, pelo que este valor representa o volume total de altas hospitalares e não a quantidade de pessoas distintas internadas.

Hipóteses relacionadas ao período da pandemia

- O elevado volume de atendimento pode ter levado a registos incompletos ou atrasados;
- Os internamentos agendados foram reduzidos ou adiados durante a pandemia;
- Apenas os casos urgentes continuaram a ser internados de forma regular;
- A procura por cuidados hospitalares diminuiu, com menos pessoas a recorrerem aos serviços por sintomas leves.

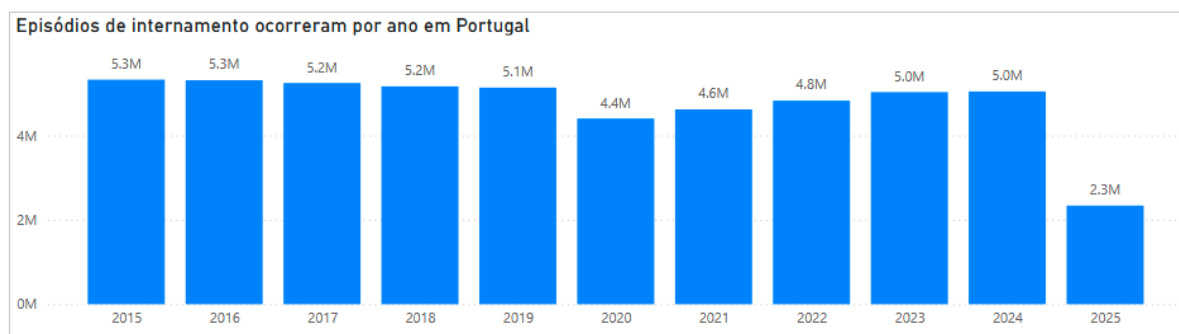


Figura 19. Evolução Anual dos Episódios de Internamento em Portugal.

```

1 Total Doentes Saídos =
2 SUM ( 'atividade-de-internamento-hospitalar'[Doentes Saídos] )

```

Figura 20. Medida DAX para Cálculo do Total de Doentes Saídos.

```

1 Dim_Calendarario =
2 CALENDARAUTO()

```

Figura 21. Criação da Tabela Dim_Calendarario com DAX.

Qual é o tempo médio de internamento por especialidade e quais áreas apresentam maior permanência?

A análise da média de dias de internamento por especialidade médica permite compreender as diferenças de complexidade entre os tipos de atendimento hospitalar ao longo dos anos. Os dados mostram que as especialidades apresentam padrões bastante distintos: os internamentos cirúrgicos tendem a ser os mais curtos, com uma média próxima de 6 dias, refletindo procedimentos programados e recuperação relativamente rápida; já a especialidade médica apresenta médias superiores, em torno de 9 a 10 dias, associadas a casos clínicos mais longos e de maior variabilidade; por fim, o grupo “Outras Camas” possui as maiores médias, próximas de 14 a 15 dias, sugerindo episódios de internamento prolongado, muitas vezes relacionados a cuidados continuados, reabilitação ou situações de longa permanência.

Ao analisar a evolução ao longo dos anos, observa-se uma tendência relativamente estável dentro de cada especialidade, sem variações expressivas que indiquem mudanças estruturais

no perfil dos internamentos. É importante destacar que este *dataset* apresenta valores nulos tanto em Doentes Saídos quanto em Dias de Internamento, especialmente em algumas especialidades. Para não distorcer a média, optámos por não substituir nulos por zero, pois zero representa uma informação válida (“houve internamento zero dias”), enquanto o nulo representa ausência de registo. Dessa forma, os valores em falta foram excluídos do cálculo, garantindo que as médias aqui apresentadas reflitam apenas registos completos e fiáveis. A presença desses nulos é monitorizada através das colunas auxiliares criadas no ETL, permitindo identificar especialidades com maior fragilidade na qualidade da informação.

Em conjunto, esta análise permite concluir que a duração média dos internamentos segue padrões clínicos previsíveis, com diferenças claras entre especialidades e grande estabilidade ao longo dos anos. Isso sugere que, apesar das oscilações no volume total de internamentos (particularmente durante a pandemia), o perfil de permanência dos doentes dentro de cada especialidade manteve-se consistente, o que é um sinal de previsibilidade operacional para a gestão hospitalar.

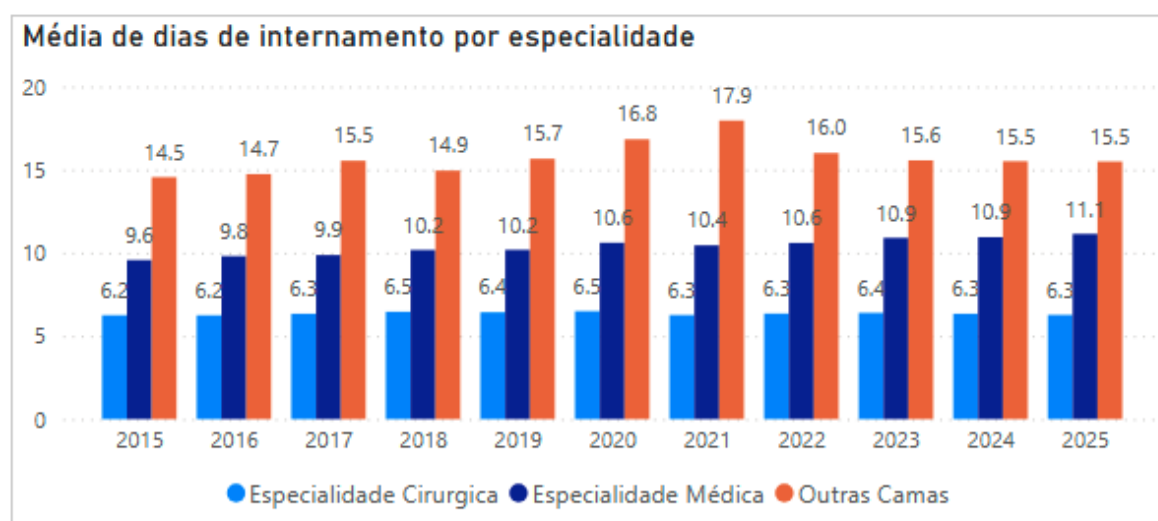


Figura 22. Média de Dias de Internamento por Especialidade (2015–2025).

```

1 Media Dias por Especialidade =
2 DIVIDE(
3     [Total Dias Internamento],
4     [Total Doentes Saídos]
5 )
6

```

Figura 23. Medida DAX para Cálculo da Média de Dias de Internamento por Especialidade.

```

1 Dim_Especialidade =
2 DISTINCT (
3     'atividade-de-internamento-hospitalar'[Tipo de Especialidade]
4 )

```

Figura 24. Criação da Tabela de Dimensão “Especialidade” com DAX.

Como estão distribuídos os internamentos pelas diferentes especialidades hospitalares?

A análise mostra que a maioria dos internamentos está concentrada em duas áreas principais. A Especialidade Cirúrgica representa cerca de 50% do total, seguida pela Especialidade Médica, com aproximadamente 46%. As Outras Camas correspondem a uma pequena fração, cerca de 4% dos episódios.

Esta distribuição indica que, no conjunto de hospitais analisados, quase todos os internamentos pertencem aos cuidados médicos e cirúrgicos tradicionais, refletindo a estrutura típica da prestação hospitalar em Portugal.

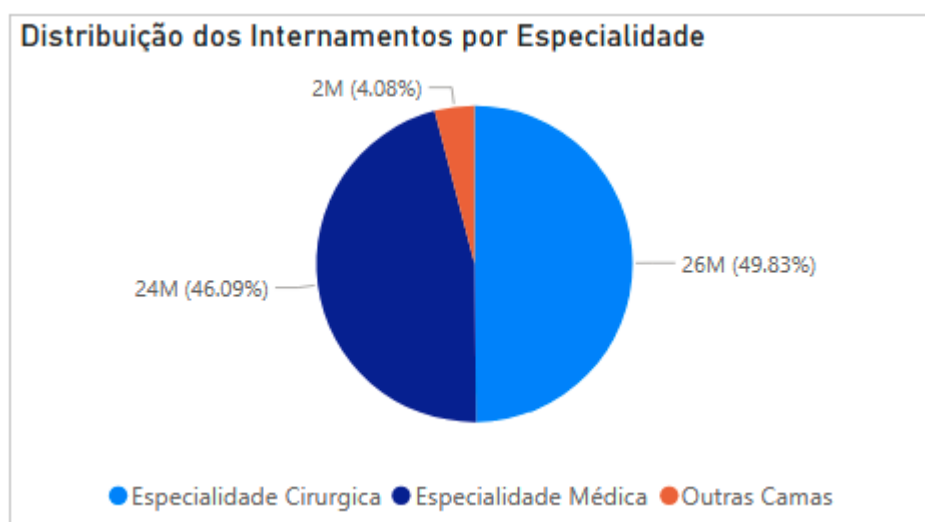


Figura 25. Distribuição dos Internamentos por Tipo de Especialidade.

```

1 Total de Doentes Saídos =
2 SUM('atividade-de-internamento-hospitalar'[Doentes Saídos])
3

```

Figura 26. Medida DAX: Total de Doentes Saídos.

```

1 Dim_Especialidade =
2 DISTINCT (
3   'atividade-de-internamento-hospitalar'[Tipo de Especialidade]
4 )
5

```

Figura 27. Criação da Dimensão “Especialidade”.

Quais regiões apresentam maior pressão hospitalar, considerando simultaneamente o volume de pacientes internados e a utilização das camas?

Antes de analisarmos a pressão hospitalar por região, é importante esclarecer exatamente o que cada indicador desta tabela representa, para evitar interpretações equivocadas.

O primeiro indicador é ‘Episódios de Internamento Concluídos’, que corresponde ao número total de altas hospitalares registadas no período. É importante destacar que este valor não representa pessoas únicas, mas sim episódios. Um mesmo paciente pode ter sido internado mais do que uma vez e, por isso, aparece repetido na contagem. Portanto, quando observamos que a região de Lisboa e Vale do Tejo registou cerca de 19 milhões, estamos a falar de 19 milhões de episódios de internamento concluídos ao longo de vários anos e em todos os hospitais da região.

O segundo indicador é ‘Dias de Cama Ocupada’, também chamado de cama-dia. Aqui não estamos a contar camas físicas, mas sim o total de dias em que camas estiveram ocupadas. Por exemplo: uma cama ocupada durante cinco dias corresponde a cinco cama-dia; cem camas ocupadas durante um ano somam dezenas de milhares de cama-dia. Assim, quando vemos que LVT acumulou cerca de 1 milhão de cama-dia no período, isso significa que, somando todos os hospitais e todos os anos analisados, as camas estiveram ocupadas durante aproximadamente um milhão de dias no total.

Com base nesses dois indicadores, o volume de internamentos e a carga de dias de cama ocupada, construímos um Índice de Pressão Hospitalar. Este índice não é um valor clínico absoluto, mas sim um indicador relativo criado para comparar as regiões entre si. Ele varia entre 0 e 2, onde valores mais próximos de 2 indicam maior pressão relativa dentro do conjunto analisado.

O que observamos é que Lisboa e Vale do Tejo, com um índice de 1.95, é claramente a região mais pressionada. Isto significa que, dentro do grupo, é a região que combina

simultaneamente o maior volume de internamentos concluídos e uma utilização muito elevada dos dias de cama. A região Norte apresenta um padrão muito semelhante, com pressão também elevada. Já Centro apresenta pressão moderada, enquanto Algarve e Alentejo registam níveis significativamente mais baixos.

Em resumo: esta tabela mostra não apenas a quantidade de atividade hospitalar, mas como essa atividade se traduz em pressão sobre a capacidade das regiões. Lisboa e Vale do Tejo e Norte destacam-se como as regiões que operam com maior intensidade relativa.

Região	Internamentos Concluídos	Dias de Cama Ocupada	Índice Pressão Hospitalar
Região de Saúde LVT	19M	1M	1.95
Região de Saúde Norte	20M	1M	1.92
Região de Saúde do Centro	10M	690K	1.01
Região de Saúde do Alentejo	2M	133K	0.01
Região de Saúde do Algarve	2M	139K	0.01
Total	52M	3M	5.86

Figura 28. Comparação da Pressão Hospitalar por Região de Saúde.

As regiões de saúde em Portugal apresentam dimensões e realidades muito diferentes. Regiões como LVT e Norte concentram a maior parte da população e, consequentemente, registam valores muito elevados tanto no número de doentes internados como nos dias de cama ocupada. Já regiões menores, como Algarve e Alentejo, apresentam volumes naturalmente inferiores.

Se utilizássemos estes valores absolutos para avaliar a pressão hospitalar, as regiões maiores pareceriam sempre mais pressionadas, mesmo que estivessem a operar de forma mais eficiente. Para garantir uma comparação justa entre regiões de tamanhos tão distintos, foi aplicada uma normalização Min–Max aos dois indicadores fundamentais: (1) total de doentes internados e (2) total de dias de cama ocupada.

A normalização Min–Max transforma cada indicador numa escala comum entre 0 e 1, onde:

- 0 representa a região com o menor valor observado,
- 1 representa a região com o maior valor,
- e os valores intermédios posicionam proporcionalmente as demais regiões.

Desta forma, tanto o volume de pacientes quanto a utilização de camas deixam de ser influenciados pela dimensão populacional ou capacidade instalada de cada região. O resultado

final é uma medida de pressão hospitalar relativa, que reflete de forma mais fiel quão sobrecarregada está uma região em comparação às outras, considerando simultaneamente a procura (internamentos) e a oferta utilizada (camas ocupadas).

```
1 Pacientes Normalizado =
2 VAR MinPac =
3     MINX(
4         ADDCOLUMNS(
5             ALL(Dim_Regiao),
6             "@Pac", [Total de Doentes Saidos]
7         ),
8         [@Pac]
9     )
10 VAR MaxPac =
11     MAXX(
12         ADDCOLUMNS(
13             ALL(Dim_Regiao),
14             "@Pac", [Total de Doentes Saidos]
15         ),
16         [@Pac]
17     )
18 RETURN
19 DIVIDE(
20     [Total de Doentes Saidos] - MinPac,
21     MaxPac - MinPac
22 )
```

Figura 29. Cálculo DAX para Normalização do Total de Doentes Saídos por Região.

```
1 Lotacao Normalizada =
2 VAR MinLot =
3     MINX(
4         ADDCOLUMNS(
5             ALL(Dim_Regiao),
6             "@Lot", [Total Lotacao Praticada]
7         ),
8         [@Lot]
9     )
10 VAR MaxLot =
11     MAXX(
12         ADDCOLUMNS(
13             ALL(Dim_Regiao),
14             "@Lot", [Total Lotacao Praticada]
15         ),
16         [@Lot]
17     )
18 RETURN
19 DIVIDE(
20     [Total Lotacao Praticada] - MinLot,
21     MaxLot - MinLot
22 )
```

Figura 30. Cálculo DAX para Normalização da Lotação Praticada por Região.

```

1 Indice Pressao Hospitalar (Fmt) =
2 FORMAT([Indice Pressao Hospitalar], "0.00")
3

```

Figura 31. Medida DAX para Formatação do Índice de Pressão Hospitalar.

```

1 Total Doentes Saídos (Fmt) =
2 VAR v = [Total de Doentes Saídos]
3 RETURN
4 SWITCH(
5     TRUE(),
6     v >= 1000000, FORMAT(ROUND(v/1000000,0), "0") & "M",
7     v >= 1000,    FORMAT(ROUND(v/1000,0), "0") & "K",
8     | | | | |    FORMAT(v,"0")
9 )
10

```

Figura 32. Medida DAX para Formatação do Total de Doentes Saídos.

```

1 Total Lotacao Praticada (Fmt) =
2 VAR v = [Total Lotacao Praticada]
3 RETURN
4 SWITCH(
5     TRUE(),
6     v >= 1000000, FORMAT(ROUND(v/1000000,0), "0") & "M",
7     v >= 1000,    FORMAT(ROUND(v/1000,0), "0") & "K",
8     | | | | |    FORMAT(v,"0")
9 )
10

```

Figura 33. Medida DAX para Formatação da Lotação Praticada.

```

1 Dim_Regiao =
2 DISTINCT (
3     UNION (
4         SELECTCOLUMNS (
5             'atividade-de-internamento-hospitalar',
6             "Regiao", 'atividade-de-internamento-hospitalar'[Região]
7         ),
8         SELECTCOLUMNS (
9             'ocupacao-do-internamento',
10            "Regiao", 'ocupacao-do-internamento'[Região]
11        )
12    )
13 )

```

Figura 34. Criação da Tabela Dimensão Região (DAX).

Qual é a evolução da carga assistencial ao longo dos anos (internamentos × dias de cama ocupada)?

O gráfico mostra que, entre 2013 e 2024, os internamentos concluídos e os dias de cama ocupada mantiveram um comportamento estável, com oscilações pequenas e esperadas na atividade hospitalar. Os internamentos situam-se entre 4 e 6 milhões por ano, enquanto a ocupação se mantém entre 200 e 300 mil cama-dia anuais, valores consistentes com a operação normal dos hospitais do SNS.

Nota: o indicador “cama-dia” não representa o número total de camas físicas, mas sim o total acumulado de camas ocupadas ao longo do ano. Valores entre 200 e 300 mil cama-dia correspondem, em média, a 500 a 800 camas ocupadas diariamente, o que está alinhado com a dimensão da rede hospitalar portuguesa.

Em 2025 observa-se uma queda acentuada tanto nos internamentos como na ocupação. Esse valor não indica uma redução real da atividade, mas sim o facto de o *dataset* conter apenas dados parciais do ano.

Importância desta análise: compreender a evolução da carga assistencial é essencial porque revela como os hospitais estão a ser utilizados ao longo do tempo. Esta informação ajuda a identificar períodos de maior pressão, compreender o impacto de acontecimentos excecionais (como a pandemia) e apoiar decisões estratégicas sobre alocação de camas, gestão de recursos humanos e planeamento da capacidade. Uma carga assistencial estável indica previsibilidade na operação hospitalar, enquanto variações abruptas podem sinalizar problemas, necessidades adicionais ou alterações nos padrões de procura.



Figura 35. Evolução da Carga Assistencial Hospitalar (2014–2025).

```
1 Total de Doentes Saídos =  
2 SUM('atividade-de-internamento-hospitalar'[Doentes Saídos])  
3
```

Figura 36. Código DAX para Cálculo do Total de Doentes Saídos.

```

1 Total Lotacao Praticada =
2 SUM('ocupacao-do-internamento'[Lotação Praticada])
3

```

Figura 37. Código DAX para Cálculo da Lotação Praticada.

```

1 Dim_Calendarario =
2 CALENDARAUTO()

```

Figura 38. Código DAX para Criação da Tabela de Datas.

Qual a participação de cada região na ocupação total de camas entre 2013 e 2025?

A análise da participação das regiões na ocupação hospitalar mostra como cada parte do país contribui para o total de camas ocupadas nos hospitais ao longo dos anos. Estas percentagens representam a fatia que cada região tem do total nacional de dias de cama ocupada, e não o número absoluto de internamentos. Ou seja, indicam qual parcela da ocupação hospitalar é gerada por cada região, refletindo o peso relativo de cada uma no sistema de saúde.

Os resultados indicam uma distribuição bastante estável ao longo do tempo. Lisboa e Vale do Tejo (36%) e a Região Norte (34%) concentram quase 70% de toda a ocupação hospitalar do país. Isso acontece porque são as regiões mais populosas e com maior número de hospitais, naturalmente gerando mais episódios de internamento e maior utilização de camas.

A Região Centro contribui com cerca de 21%, mantendo um peso intermédio, enquanto Algarve e Alentejo representam as menores parcelas, ambas próximas de 4%. Esses valores são coerentes com a menor população dessas regiões e com uma rede hospitalar mais reduzida.

Mesmo durante os anos da pandemia (2020–2022), quando houve queda geral de internamentos, a distribuição percentual manteve-se praticamente igual. Isso significa que todas as regiões foram afetadas de forma proporcional, sem alterações relevantes no equilíbrio nacional.

Esta análise é importante porque confirma que a ocupação hospitalar segue muito de perto a distribuição populacional e a capacidade instalada, sem desvios inesperados ao longo dos anos. Também ajuda a gestão a perceber onde a pressão hospitalar é naturalmente maior e onde deve haver maior investimento em capacidade, recursos e planeamento estratégico.

Regiao	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	Total
Região de Saúde LVT	2.95%	2.83%	2.91%	2.92%	2.97%	2.95%	2.91%	2.86%	2.84%	2.82%	2.69%	2.80%	1.87%	36.30%
Região de Saúde Norte	2.49%	2.56%	2.57%	2.57%	2.59%	2.65%	2.60%	2.70%	2.72%	2.76%	2.76%	2.80%	1.89%	33.66%
Região de Saúde do Centro	1.82%	1.78%	1.76%	1.75%	1.77%	1.72%	1.70%	1.67%	1.67%	1.64%	1.63%	1.59%	1.04%	21.53%
Região de Saúde do Algarve	0.33%	0.33%	0.34%	0.34%	0.34%	0.35%	0.35%	0.34%	0.37%	0.35%	0.34%	0.34%	0.23%	4.35%
Região de Saúde do Alentejo	0.35%	0.33%	0.33%	0.33%	0.33%	0.32%	0.32%	0.32%	0.33%	0.33%	0.32%	0.32%	0.22%	4.15%
Total	7.94%	7.81%	7.91%	7.91%	8.00%	7.99%	7.88%	7.91%	7.93%	7.90%	7.74%	7.84%	5.24%	100.00%

Figura 39. Participação das Regiões na Ocupação Hospitalar ao Longo dos Anos.

```

1 Total Lotacao Praticada =
2 SUM('ocupacao-do-internamento'[Lotação Praticada])
3

```

Figura 40. Código DAX para Cálculo do Total da Lotação Praticada.

```

1 Dim_Regiao =
2 DISTINCT (
3     UNION (
4         SELECTCOLUMNS (
5             'atividade-de-internamento-hospitalar',
6             "Regiao", 'atividade-de-internamento-hospitalar'[Região]
7         ),
8         SELECTCOLUMNS (
9             'ocupacao-do-internamento',
10            "Regiao", 'ocupacao-do-internamento'[Região]
11        )
12    )
13 )

```

Figura 41. Código DAX para Criação da Tabela Dimensional de Regiões.

```

1 Dim_Calendarario =
2 CALENDARAUTO()

```

Figura 42. Código DAX para Criação Automática da Tabela de Datas.

Qual especialidade é responsável pela maior carga assistencial (total de dias de internamento) nos últimos três anos?

Para responder a esta pergunta, criámos uma medida em DAX que calcula o total de dias de internamento considerando apenas os três anos mais recentes. A partir desse cálculo, observamos que a Especialidade Médica apresenta a maior carga assistencial, acumulando mais dias de internamento do que as demais especialidades. Isso sugere que os internamentos clínicos tendem a ser mais prolongados e exigem maior utilização das camas.

A Especialidade Cirúrgica aparece em segundo lugar, com internamentos mais frequentes, porém mais curtos. Já o grupo de Outras Camas tem menos episódios, mas inclui casos de longa permanência, o que também impacta o total de dias.

Importante destacar que estas conclusões são interpretações baseadas nos dados disponíveis e devem ser encaradas como hipóteses, podendo variar conforme contexto clínico, perfil populacional e práticas assistenciais de cada instituição.

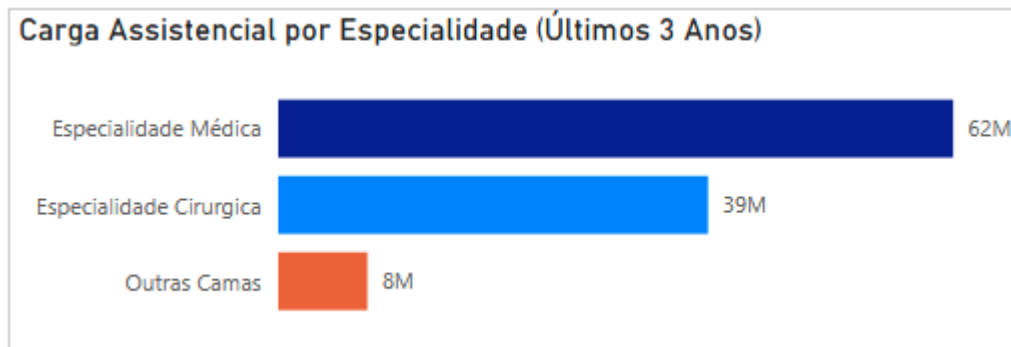


Figura 43. Carga Assistencial por Especialidade nos Últimos 3 Anos.

```

1 Total Dias Internamento Últimos 3 Anos =
2 VAR MaxAno = CALCULATE(MAX('Dim_Calendario'[Ano]), ALL('Dim_Calendario'))
3 VAR Inicio = MaxAno - 2
4 RETURN
5 CALCULATE(
6     [Total Dias Internamento],
7     'Dim_Calendario'[Ano] >= Inicio
8 )
9

```

Figura 44. Código DAX para Cálculo dos Dias de Internamento dos Últimos 3 Anos.

```

1 Dim_Especialidade =
2 DISTINCT (
3     'atividade-de-internamento-hospitalar'[Tipo de Especialidade]
4 )
5

```

Figura 45. Código DAX para Criação da Tabela de Especialidades.

Qual será a evolução do número de doentes internados nos próximos períodos?

A análise da tendência histórica dos internamentos, juntamente com a previsão automática do *Power BI*, mostra que o número de doentes internados deverá voltar a subir após a queda de 2025 (ano com dados incompletos). Para 2026, o modelo estima um valor próximo dos anos anteriores, sugerindo recuperação da atividade.

Nos anos seguintes, a previsão indica que os internamentos devem manter-se relativamente estáveis, sem grandes aumentos ou quedas. O intervalo sombreado mostra possíveis variações, mas, no geral, espera-se uma evolução consistente e sem mudanças abruptas.

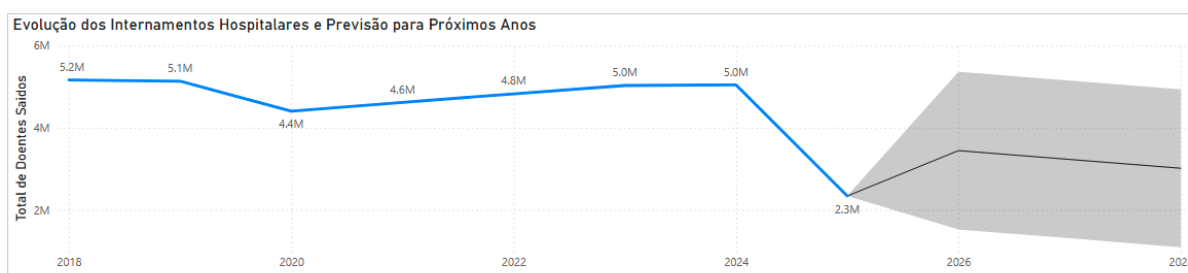


Figura 46. Evolução dos Internamentos Hospitalares e Previsão para os Próximos Anos.

```
1 Total de Doentes Saídos =  
2 SUM('atividade-de-internamento-hospitalar'[Doentes Saídos])  
3
```

Figura 47. Código DAX para o Cálculo do Total de Doentes Saídos.

```
1 Dim_Calendarario =  
2 CALENDAR(AUTO())
```

Figura 48. Código DAX para Criação da Tabela de Calendário.

Qual será a evolução da Lotação praticada nos próximos períodos?

A análise mostra que a lotação praticada (média diária de camas ocupadas) manteve-se estável entre 2018 e 2024, com valores próximos de 250 mil cama-dia por ano. Em 2025 ocorre uma queda acentuada, que provavelmente se deve a dados incompletos no *dataset*.

A previsão indica que, a partir de 2026, a lotação volta a aproximar-se do padrão dos anos anteriores, sugerindo recuperação da ocupação hospitalar. Nos períodos seguintes, o modelo aponta para uma evolução relativamente estável, sem aumentos bruscos. A área sombreada representa possíveis variações, naturais em previsões de longo prazo, mas reforça que a tendência geral permanece estável.

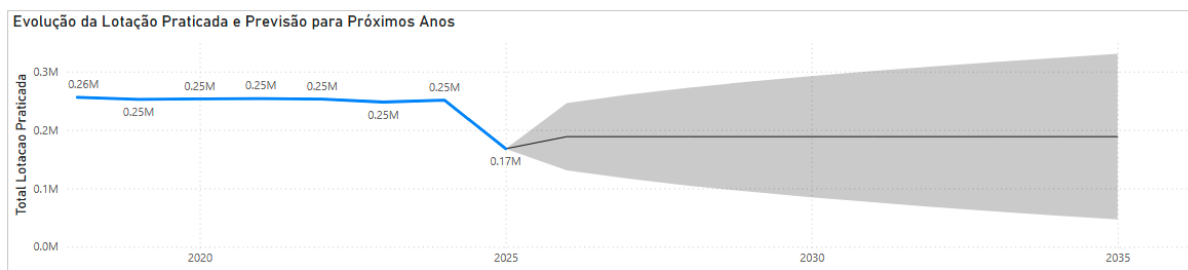


Figura 49. Evolução da Lotação Praticada e Previsão para os Próximos Anos.

```

1 Total Lotacao Praticada =
2 SUM('ocupacao-do-internamento'[Lotação Praticada])
3

```

Figura 50. Código DAX para Cálculo do Total de Lotação Praticada.

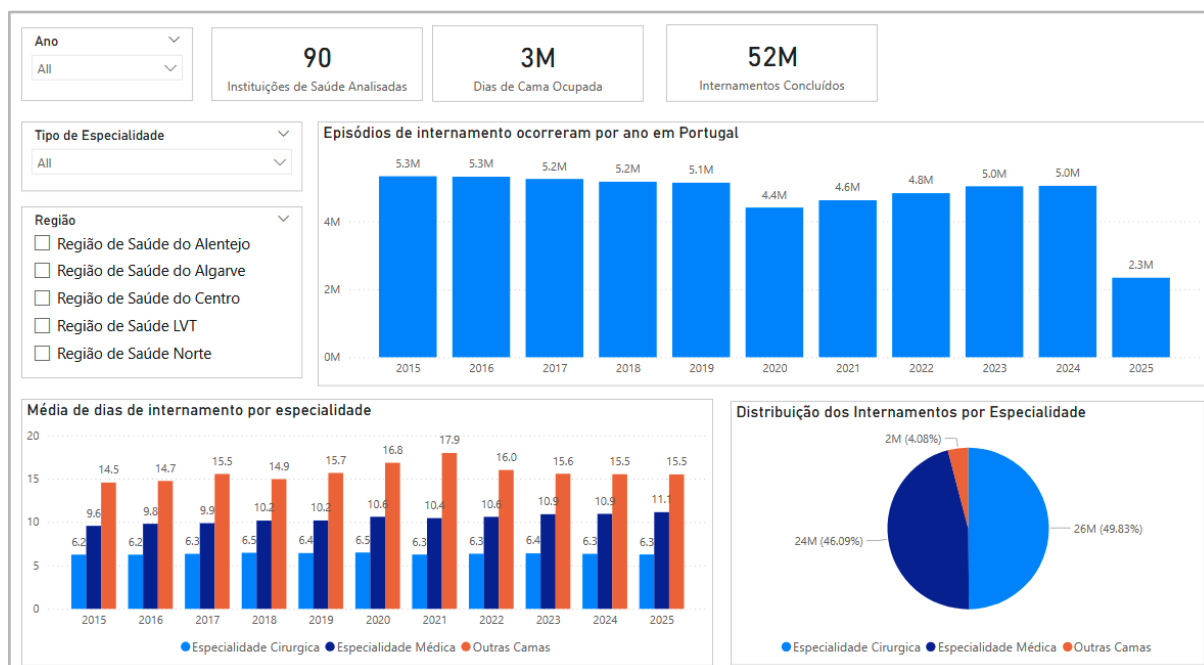
```

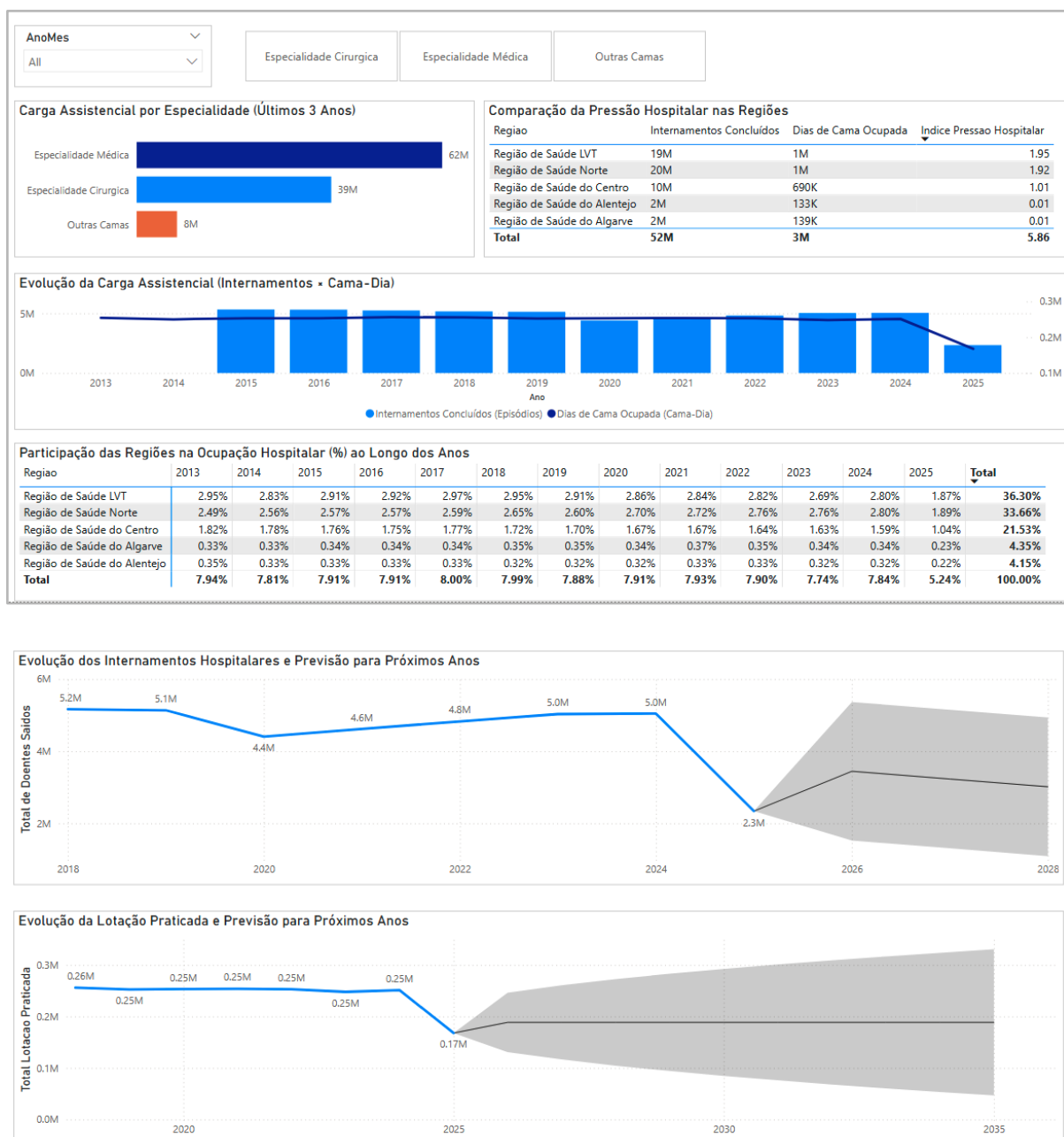
1 Dim_Calendarario =
2 CALENDAR(AUTO())

```

Figura 51. Código DAX para Criação Automática da Tabela de Calendário.

Visualização do Power BI completo:





7. DISCUSSÃO DE RESULTADOS E PROPOSTA DE TRABALHO FUTURO

Os resultados obtidos neste trabalho permitiram uma visão abrangente sobre a evolução da atividade hospitalar em Portugal ao longo da última década. Através da análise dos dados de internamentos hospitalares, foi possível verificar que o sistema de saúde português apresenta um volume anual elevado e relativamente estável de episódios de internamento, com uma redução significativa observada no ano de 2020, reflexo direto do impacto da pandemia de COVID-19. Essa quebra foi, no entanto, seguida de uma recuperação progressiva nos anos seguintes, o que indica uma resiliência do sistema hospitalar perante eventos de grande escala.

As previsões geradas para os anos futuros sugerem a continuidade desta tendência histórica, assumindo que, na ausência de choques externos ou alterações estruturais significativas, o número de internamentos deverá manter-se dentro de valores esperados. Esta estabilidade permite antecipar necessidades em termos de recursos humanos, capacidade instalada e gestão hospitalar, e pode apoiar políticas públicas mais informadas.

Paralelamente, a análise da lotação praticada (medida em dias de cama ocupada) revelou igualmente um comportamento estável ao longo dos anos, refletindo uma gestão consistente da capacidade hospitalar. A ausência de grandes variações neste indicador sugere que a relação entre oferta e procura de camas hospitalares tem sido, de um modo geral, equilibrada. As projeções apontam para uma manutenção deste padrão, o que reforça a noção de controlo da pressão hospitalar, pelo menos no cenário analisado.

Contudo, é importante reconhecer as limitações da abordagem adotada. Em primeiro lugar, destaca-se a limitação dos dados, nomeadamente a informação incompleta relativa ao ano de 2025, o que compromete, ainda que parcialmente, a fiabilidade das previsões a curto prazo. Além disso, o modelo preditivo utilizado baseou-se exclusivamente em regressão linear, uma técnica estatística simples que não incorpora sazonalidade, variações mensais ou relações com variáveis externas – fatores frequentemente relevantes na dinâmica hospitalar.

A granularidade dos dados constitui outra limitação significativa. A análise foi realizada com base em dados anuais agregados, o que pode ocultar variações importantes ao longo do ano ou entre diferentes serviços hospitalares. Da mesma forma, o modelo assume que as tendências passadas se manterão no futuro, o que pode ser um pressuposto frágil, especialmente perante contextos de crise, reformas no sistema de saúde ou alterações demográficas significativas.

Com base nestas limitações, delineiam-se algumas propostas para trabalhos futuros que permitam aprofundar a análise e aumentar a robustez das conclusões:

- Integração de novas variáveis: A inclusão de fatores como número de profissionais de saúde, taxas de ocupação por especialidade, tipologia dos internamentos (agudos, crónicos, cirúrgicos, etc.) e indicadores de complexidade clínica poderá enriquecer significativamente os modelos de análise. Dados demográficos, como envelhecimento da população ou distribuição geográfica, também devem ser considerados;
- Aumento da granularidade temporal e temática: A utilização de séries mensais ou trimestrais permitiria captar variações sazonais e tendências de curto prazo, com

impacto direto na precisão das previsões. Além disso, a análise segmentada por hospital, especialidade médica ou grupo de patologias pode revelar padrões até agora não identificados;

- Adoção de modelos preditivos mais avançados: Modelos como ARIMA, *Prophet*, redes neurais ou outros algoritmos de *machine learning* podem ser testados e comparados com a regressão linear utilizada neste trabalho, oferecendo maior flexibilidade e capacidade de adaptação a dados complexos.
- Cruzamento com outras bases de dados: A integração de fontes externas como o INE, SNS, dados de mortalidade, estatísticas de cuidados primários ou registos de urgência permitiria uma análise mais completa da procura de cuidados hospitalares face à capacidade de resposta do sistema.
- Desenvolvimento de *dashboards* operacionais em tempo real: A criação de painéis interativos para monitorização contínua da ocupação hospitalar, com alertas automáticos para situações de risco de saturação, pode constituir uma ferramenta de grande utilidade para a gestão operacional, especialmente em contextos de elevada pressão como surtos epidémicos ou picos sazonais de procura.

Em suma, os resultados obtidos constituem um ponto de partida sólido para compreender a atividade hospitalar recente e planear estratégias de melhoria da eficiência e sustentabilidade do sistema de saúde. Contudo, para uma abordagem mais robusta e preditiva, será essencial integrar mais variáveis, aumentar a frequência e o detalhe dos dados analisados, e aplicar metodologias analíticas mais sofisticadas que reflitam a complexidade real do sistema de saúde português.

CONCLUSÕES

O presente estudo permitiu uma análise abrangente da atividade hospitalar no Serviço Nacional de Saúde em Portugal entre 2013 e 2025, focando-se nos internamentos e na ocupação hospitalar. Através da utilização de ferramentas de *Business Intelligence*, como o *Power BI* e *Power Query*, foi possível construir um modelo analítico robusto e acessível, fundamentado em dados públicos e metodologias transparentes.

Verificou-se que o sistema hospitalar nacional apresenta um padrão estável de internamentos e de ocupação de camas, com uma quebra significativa em 2020, associada à pandemia de COVID-19, seguida de uma recuperação progressiva. As projeções geradas por modelos de regressão linear apontam para a manutenção destas tendências, oferecendo uma base sólida para o planeamento hospitalar futuro.

Apesar dos resultados promissores, o trabalho enfrenta limitações importantes, como a utilização de dados anuais agregados, a simplicidade dos modelos preditivos aplicados e a ausência de variáveis explicativas adicionais. Esses fatores restringem a capacidade de previsão em contextos de maior complexidade ou variação súbita na procura de cuidados.

Com base nestes pontos, propõem-se diversos caminhos para trabalhos futuros, nomeadamente a integração de dados de maior granularidade temporal, o cruzamento com outras fontes externas e a adoção de modelos mais sofisticados de previsão. Estas melhorias permitirão não apenas um maior rigor nas análises, mas também a geração de insights mais acionáveis para apoiar a gestão em saúde.

O estudo evidencia, assim, o valor das abordagens de *Business Intelligence* no setor público da saúde, demonstrando que é possível, com dados acessíveis e ferramentas adequadas, produzir conhecimento relevante e orientado para a melhoria contínua dos serviços hospitalares.

BIBLIOGRAFIA

- Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *Journal of big data*, 9(1), 3. <https://doi.org/10.1186/s40537-021-00553-4>
- Gomes, M. A. S. (2022). Análise de big data aplicada a serviços de saúde. *Conference Paper*. “Harnessing Big Data and Predictive Analytics for Modern Healthcare Management.” (2025). *IIARD Journals*. <https://www.iiardjournals.org>
- Nwoke, J. (2024). Healthcare Data Analytics and Predictive Modelling: Enhancing Outcomes in Resource Allocation, Disease Prevalence and High-Risk Populations. *International Journal of Health Sciences*, 7(7), 1–35.
- Worku, A. (2025). Introduction to healthcare data analytics – An overview. *Health Informatics Review*. SpringerLink.
- Worku, A., Arage, F.G. & Kebede, F.B. Introduction to health care data analytics- an overview. *Discov Health Systems* 4, 107 (2025). <https://doi.org/10.1007/s44250-025-00291-x>