

## **Fundamentação dos dados**

**De onde vêm os dados? Como “recolheram” os dados? A fonte é confiável? Há alguma outra consideração que tiveram em consideração ao “recolher” estes dados?**

**Fonte dos dados:** SNS - Serviço Nacional de Saúde, através dos relatórios do Morbilidade e Mortalidade Hospitalar, da base de dados da SNS - Portal da Transparência, e em Dados.Gov.Pt em 2021. <https://transparencia.sns.gov.pt/explore/?sort=modified>. Todavia recentemente foi retirado da plataforma este conjunto de dados e como tal atualmente o ficheiro não se encontra disponível.

Durante o processo de pesquisa, inicialmente, encontramos este tema na plataforma de Dados.Gov.Pt. Todavia, esse mesmo website dava indicações de onde provinha a fonte, que no caso seria a base de dados do Serviço Nacional de Saúde (SNS) – Portal da Transparência. De tal modo, fomos averiguar e após uma discussão de ideias optámos por utilizar apenas os dados do ano de 2021 e recolher os dados estatísticos de apenas 3 tipos de causas de morte, uma vez que, as que eram apresentadas eram número bastante numeroso.

Na escolha das causas de mortalidade decidimos escolher dois géneros de morbilidade que são acontecem mais frequentemente e um menos frequente. Outro fator que também influenciou as nossas escolhas, foi o facto de que o ano de 2021 ter sido marcado pelo número de mortalidade devido ao Covid-19 e como tal optámos por escolher causas de morte que também fossem mais a esse encontro.

**“Limpeza” os dados? Existem valores “ausentes”? Eles ocorrem em campos que são importantes para os objetivos do projeto? Existem duplicados? Eles ocorrem em campos que são importantes para os objetivos do projeto?**

Felizmente no nosso projeto, não tivemos problemas com dados ausentes, duplicados ou nada que possa parecer.

**Quais são os campos mais relevantes para os objetivos do projeto?**

Para este projeto os campos que a nosso ver eram mais importantes e como tal foram também os que mais usamos foram: o campo do período, o tipo de doença, o número de óbitos, a faixa etária e o género.

**Há algum problema de tipo de dados (por exemplo, palavras em campos que deveriam ser numéricos, datas, horas...)? Como serão corrigidos?**

Neste trabalho durante a conceção do mesmo, encontramos um erro na construção de um gráfico, todavia, esse erro era do próprio ficheiro .csv e não havia forma de resolvê-lo.

No código é possível ver que tentou-se usar diferentes funções para tentar entender se o erro era nosso ou do próprio ficheiro .csv.

## **Precisam de rejeitar algum dado? Quais dados? Porquê? Há alguma razão para que isso possa afetar as análises que pretendem fazer ou as conclusões poderão chegar?**

Durante o desenvolvimento do trabalho percebemos que não havia necessidade de usar todas as estatísticas que nos eram fornecidas e como tal escolhemos logo no início eliminar essas estatísticas que não íamos usar e que ao mesmo tempo não eram influenciáveis para a interpretação dos dados, como é o caso da coluna “Código do Capítulo Diagnóstico ICD9CM/ICD10CMPCS “. O objetivo desta coluna é catalogizar o tipo de doença por números, o que seria uma mais-valia se tivéssemos todas as causas de morte juntas só num ficheiro .csv, mas como não é o caso, eliminámos.

Neste projeto fizemos uma alteração nos títulos das seguintes colunas: “Descrição do Capítulo Diagnóstico ICD9CM/ICD10CMPCS”. Como podemos ver este título não era prático nem apropriado para ser utilizado sempre que queríamos referi-lo, por isso optámos por substituir o nome para algo mais simples como: “Descrição da Doença”.

### **Contexto**

Em Portugal, existem vários órgãos que anualmente abordam as diferentes vertentes da saúde (ex: perfil de saúde da população, acesso, resultados de saúde, financiamento). São exemplos a Direção-Geral da Saúde, o Instituto Nacional de Saúde e o Serviço Nacional de Saúde. Eles analisam, em profundidade e com dados atualizados organização e governança, financiamento, recursos físicos e humanos, prestação de serviços e prestação de cuidados.

Este trabalho tem como ponto de partida os relatórios e documentos que são produzidos em Portugal e que analisam total ou parcialmente questões relacionadas à situação da saúde no país.

Neste sentido, este projeto dedica-se a compreender quais foram as principais causas de mortalidade e morbilidade em Portugal no ano de 2021.

### **Estrutura**

**aparelho\_circulatorio.csv**: Contém dados mensais relativos às doenças do aparelho circulatório, extraídos do painel Portal da Transparência do SNS. Esta série de dados tem início a 01-2021 até 12-2021.

**aparelho\_respiratorio.csv**: Contém dados mensais relativos às doenças do aparelho respiratório, extraídos do painel Portal da Transparência do SNS. Esta série de dados tem início a 01-2021 até 12-2021.

**perturbacoes\_mentais.csv**: Contém dados mensais relativos aos transtornos mentais, comportamentais e de neuro desenvolvimento, extraídos do painel Portal da Transparência do SNS. Esta série de dados tem início a 01-2021 até 12-2021.

### **Dicionário de dados**

Apresentação do dicionário de dados sobre o `aparelho_circulatorio.csv` , o `aparelho_respiratorio.csv` e das `perturbacoes_mentais.csv`

Nome da coluna	Significado	Valores Possíveis
Período	Ano e mês da coleta dos dados apresentados.	AAAA-MM-DD
Código Capítulo Diagnóstico ICD9CM/ICD10CMPCS	Código que está associado ao tipo de doença descrito	Inteiro > = 0
Descrição Capítulo Diagnóstico ICD9CM/ICD10CMPCS	Categoria da doença descrita	Doenças do aparelho circulatório, Doenças do aparelho respiratório ou Transtornos mentais, comportamentais e de neuro desenvolvimento
Instituição	Nome das instituições hospitalares	Apresentação dos nomes das instituições hospitalares
Região	Nome da região de Portugal	Apresentação das regiões de Portugal de norte a sul do país e das regiões autónomas (Madeira e Açores)
Faixa Etária	Diferentes faixas etárias	Apresentação das faixas etárias em intervalos abertos
Género	Classe de género	Género masculino ou feminino
Internamentos	Número de internamentos	Inteiro > = 0
Dias de Internamento	Número de dias de internamento	Inteiro > = 0
Ambulatório	Número de casos ambulatoriais	Inteiro > = 0
Óbitos	Número de óbitos	Inteiro > = 0

### **Bibliografia**

<https://dados.gov.pt/pt/datasets/morbilidade-e-mortalidade-hospitalar/>

<https://transparencia.sns.gov.pt/explore/?sort=modified>