

# Visualização de Informação

## CHECKPOINT II: Limpeza e Processamento dos dados

G22 - A

### 1. *Dataset* Inicial

Para explorarmos a domínio proposto, iremos recorrer as variáveis seguintes disponíveis na base de dados estatísticos PORDATA.

- Ganho médio mensal dos trabalhadores por contra de outrem: total e por nível de escolaridade.
- Poder de compra *per capita*.
- Centros de saúde: consultas por habitante.

As mesmas serão extraídas primeiramente a nível municipal (NUTS III), passando para os totais de NUTS II (Norte, Centro, Área Metropolitana de Lisboa, Alentejo, Algarve e Regiões Autónomas da Madeira e Açores) e, por fim, NUTS I: totais de Continente e Regiões Autónomas.

Este *dataset* contém, sem qualquer tipo de tratamento, perto das 16.000 entradas e está disponível no site cito.

### 2. Dados selecionados/derivados

No site da PORDATA é possível fazer um pré-processamento dos dados de cada variável antes de descarregar um ficheiro .csv completo.

Assim, de entre todos os anos disponíveis desde 1985 para a primeira variável, ficamos apenas com o intervalo de anos entre 2009 e 2013, já que o nosso objectivo é estudar o efeito da mais recente recessão. Em relação à variável “Poder de compra per capita”, existem dados disponíveis desde 1993, mas permanecemos com o intervalo dito acima. Para a terceira variável, houve algumas alterações: nomeadamente, no primeiro *checkpoint* estava a ser tido em conta o número do consultas por habitante em centros de saúde do Serviço Nacional de Saúde. Ora, concluímos que com a esta variável, não existem valores municipais e regionais nem para a Região Autónoma da Madeira (RAM) nem para os Açores (RAA), uma vez que têm Serviços de Saúde Regionais: respectivamente SESARAM e SRSAçores. Para que pudéssemos ter mais variáveis com que trabalhar e que nos permitissem criar uma visualização mais rica, consideramos o número de consultas por habitante em unidades básicas de saúde nacionais. Contudo, apenas conseguimos aumentar o total de valores para contemplar a RAM, não existindo valores da RAA.

Adicionalmente, importámos os totais de NUTS II e NUTS I para o *dataset*, os quais não possuíamos anteriormente e que serão necessários para cumprir as tarefas propostas.

### 3. Abstração de Dados

- Tipo do *dataset*: Tabela estática.
- Descrição dos itens: Cada entrada da tabela representa uma Unidade Territorial, seja ela NUTS I, NUTS II ou NUTS III.
- Tipos e descrições dos atributos:
  - “**Âmbito Geográfico**” – Atributo categórico, identificador de cada NUT, podendo formar grupos de territórios cujo âmbito geográfico seja o mesmo.
  - “**Nome do Território.**” – Atributo categórico, identificador de cada território, único.
  - “**Ganho Mensal dos Trabalhadores Total.**” – Atributo ordenado quantitativo,

descriptor do ganho mensal em total anual, em euros, dos trabalhadores por conta de outrem de um dado território.

- **“Ganho Mensal dos Trabalhadores com Nível de Escolaridade inferior ao 1º ciclo do básico.”** – Atributo ordenado quantitativo, descriptor do ganho mensal, em euros, dos trabalhadores por conta de outrem de um dado território, com nível de escolaridade inferior ao 1º ciclo do ensino básico.
- **“Ganho Mensal dos Trabalhadores com Nível de Escolaridade equivalente ao 1º ciclo do básico.”** – Atributo ordenado quantitativo, descriptor do ganho mensal, em euros, dos trabalhadores por conta de outrem de um dado território, com nível de escolaridade equivalente ao 1º ciclo do ensino básico.
- **“Ganho Mensal dos Trabalhadores com Nível de Escolaridade equivalente ao 2º ciclo do básico.”** – Atributo ordenado quantitativo, descriptor do ganho mensal, em euros, dos trabalhadores por conta de outrem de um dado território, com nível de escolaridade equivalente ao 2º ciclo do ensino básico.
- **“Ganho Mensal dos Trabalhadores com Nível de Escolaridade equivalente ao 3º ciclo do básico.”** – Atributo ordenado quantitativo, descriptor do ganho mensal, em euros, dos trabalhadores por conta de outrem de um dado território, com nível de escolaridade equivalente ao 3º ciclo do ensino básico.
- **“Ganho Mensal dos Trabalhadores com Nível de Escolaridade equivalente ao Secundário e Pós-Secundário.”** – Atributo ordenado quantitativo, descriptor do ganho mensal, em euros, dos trabalhadores por conta de outrem de um dado território, com nível de escolaridade equivalente ao Ensino Secundário e Pós-Secundário.
- **“Ganho Mensal dos Trabalhadores com Nível de Escolaridade equivalente ao Superior.”** – Atributo ordenado quantitativo, descriptor do ganho mensal, em euros, dos trabalhadores por conta de outrem de um dado território, com nível de escolaridade equivalente ao Ensino Superior.
- **“Poder de compra *per capita*.”** – Atributo ordenado quantitativo, descriptor do poder de compra manifestado quotidianamente, em termos *per capita*, nos diferentes municípios ou regiões, em índice, com valor 100 na média do País.
- **“Consultas por Habitante.”** – Atributo ordenado quantitativo em rácio, descriptor do número de consultas médicas que tiveram lugar nas unidades de saúde básicas de um dado território, por habitante.

#### 4. Processamento do *Dataset*

A coluna "Ganho Mensal dos Trabalhadores com Nível de Escolaridade Desconhecido" foi removida, por estar bastante incompleta e por não se poder chegar a alguma conclusão com esses valores; também as colunas de 2010, 2012 e 2013 do “Poder de Compra per capita”, e 2013 das “Consultas por Habitante” foram removidas por não existirem dados disponíveis.

Alguns dados foram corrigidos, porque houve uma atualização da tabela de dados da variável “Poder de Compra *per capita*”, a 21/10/2015. Existiam também valores em falta, que foram calculados com o valor intermédio entre o ano anterior e o ano seguinte, não deturpando a tendência; para os territórios que tinham anos consecutivos em falta, foram preenchidos com o valor "-1", uma vez que é um valor fora do domínio e pode assim ser excluído da visualização, não sendo possível obter valores relevantes e úteis para uma conclusão. Os valores médios dos NUTS II e III foram calculados tendo em conta os valores dos municípios.

Assim o *dataset* fica com cerca de 13000 entradas disponíveis para a visualização.

#### 5. Mapeamento entre *Dataset* e Questões

Mostramos agora o mapeamento entre colunas do *Dataset* e as questões que foram apresentadas no primeiro *Checkpoint*.

- **Questão 1:** “Que consequências teve o último período de recessão, na evolução do rendimento, tendo em conta a qualificação?”
  - Colunas: “Ganho Mensal dos Trabalhadores com Nível de Escolaridade inferior ao 1º ciclo do básico, equivalente ao 1º ciclo do básico, ao 2º ciclo do básico, ao 3º ciclo do básico, ao Secundário e Pós-Secundário e equivalente ao Superior”, “Âmbito Geográfico”, “Nome do Território”
  - Exemplo: Para valores de 2009 e 2013.
    - Município; Arcos de Valdevez; 542.9; 596.4; 616.1; 661.2; 628.4; 693.1; 662.1; 713.3; 765.5; 811.7; 1377.0; 1359.8
- **Questão 2:** “De que forma contribui o nível de escolaridade para a obtenção de um rendimento mais elevado?”
  - Colunas: “Ganho Mensal dos Trabalhadores com Nível de Escolaridade inferior ao 1º ciclo do básico, equivalente ao 1º ciclo do básico, ao 2º ciclo do básico, ao 3º ciclo do básico, ao Secundário e Pós-Secundário e equivalente ao Superior”, “Âmbito Geográfico”, “Nome do Território”
  - Exemplo: Para valores de 2009 e 2013.
    - Município; Arcos de Valdevez; 542.9; 596.4; 616.1; 661.2; 628.4; 693.1; 662.1; 713.3; 765.5; 811.7; 1377.0; 1359.8
- **Questão 3:** “Qual é o município onde a escolaridade menos influenciou o nível do rendimento?”
  - Colunas: “Ganho Mensal dos Trabalhadores com Nível de Escolaridade inferior ao 1º ciclo do básico, equivalente ao 1º ciclo do básico, ao 2º ciclo do básico, ao 3º ciclo do básico, ao Secundário e Pós-Secundário e equivalente ao Superior”, “Âmbito Geográfico”, “Nome do Território”
  - Exemplo: Para valores de 2009 e 2013
    - Município; Arcos de Valdevez; 542.9; 596.4; 616.1; 661.2; 628.4; 693.1; 662.1; 713.3; 765.5; 811.7; 1377.0; 1359.8
- **Questão 4:** “Como se relaciona a evolução do rendimento com o poder de compra *per capita*? Poder-se-á concluir que um nível de escolaridade elevado conduz a um poder de compra mais alto?”
  - Colunas: “Ganho Mensal dos Trabalhadores com Nível de Escolaridade inferior ao 1º ciclo do básico, equivalente ao 1º ciclo do básico, ao 2º ciclo do básico, ao 3º ciclo do básico, ao Secundário e Pós-Secundário e equivalente ao Superior”, “Poder de compra *per capita*”, “Âmbito Geográfico”, “Nome do Território”
  - Exemplo: Para valores de 2009 e 2013, para os valores de Ganho, 2 e 2011 para Poder de compra *per capita*.
    - Município; Arcos de Valdevez; 542.9; 596.4; 616.1; 661.2; 628.4; 693.1; 662.1; 713.3; 765.5; 811.7; 1377.0; 1359.8; 55.76; 61.97
- **Questão 5:** “A procura dos centros de saúde aumentou com o recente período de recessão?”
  - Colunas: “Âmbito Geográfico”, “Nome do Território”, “Consultas por Habitante”
  - Exemplo: Para valores de 2009 e 2012
    - Município; Valpaços; 2.6; 1.7
- **Questão 6:** “Que diferença existe no ganho mensal no Município de Lisboa, entre o ano de 2009 e o ano de 2013?”
  - Colunas: “Âmbito Geográfico”, “Nome do Território”, “Ganho Mensal dos Trabalhadores Total”
  - Exemplo: Município; Lisboa; 1508.8; 1574.9
- **Questão 7:** “Qual foi o município que, em 2012, se destacou pela superior procura dos centros de saúde?”
  - Colunas: “Âmbito Geográfico”, “Nome do Território”, “Consultas por Habitante”

- Exemplo: Município; Lisboa; 2.2
- **Questão 8:** “O ganho mensal aumentou ou diminuiu, a nível nacional?”
  - Colunas: “Âmbito Geográfico”, “Nome do Território”, “Ganho Mensal dos Trabalhadores Total”
  - Exemplo: Para valores de 2010 e 2013
    - NUTS I; Continente; 1076.3; 1093.8
- **Questão 9:** “Será que um ganho mensal superior, implica que haja menos procura dos centros de saúde e mais procura de unidades de saúde privadas?”
  - Colunas: “Âmbito Geográfico”, “Nome do Território”, “Ganho Mensal dos Trabalhadores Total”, “Consultas por Habitante”
  - Exemplo: Para valores de 2011 e dois municípios
    - Município; Paredes de Coura; 740.0; 3.6
    - Município; Amadora; 1249.4; 2.2
- **Questão 10:** “O poder de compra está relacionado com a procura dos centros de saúde? Poder-se-á concluir que, quanto maior o poder de compra, menor a procura de consultas em centros de saúde?”
  - Colunas: “Âmbito Geográfico”, “Nome do Território”, “Poder de compra *per capita*”, “Consultas por Habitante”
  - Exemplo: Para valores de 2011
    - Município; Amadora; 105.74; 2.2
- **Questão 11:** “Quais os 3 municípios com maior poder de compra *per capita*, e os 3 municípios com menor poder de compra *per capita*, no ano de 2009?”
  - Colunas: “Âmbito Geográfico”, “Nome do Território”, “Poder de compra *per capita*”
  - Exemplo: Para dois municípios
    - Município; Loures; 121.60
    - Município; Vila Verde; 57.83