

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

Na próxima página encontra-se uma tabela com alguns valores de exemplo, que podem dar resposta, por exemplo, às questões 1, 2 e 8 (omissas aqui por questões de extensão do

documento, mas que podem ser consultadas no ficheiro do primeiro *checkpoint*). Para as questões 1 e 2, utilizaremos os atributos “Ganho mensal dos Trabalhadores Total” e todos os atributos “Ganho mensal dos Trabalhadores com Nível de Escolaridade (...)”. Para a questão 8, utilizaremos o “Ganho mensal dos Trabalhadores Total”.

Âmbito Geog Territórios		2009 - Totê 2010 - Totê 2011 - Totê 2012 - Totê 2013 - Totê		2009 - Poder dê 2011 - Poder dê 2009 - Consult 2010 - Consult 2011 - Consult 2012 - Consultas por hab.													
NUTS 2013	Portugal	918.4	950	980.8	982.5	1001.5	...	100	100	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5
NUTS I	Continente	881.8	917.8	930.1	939.9	934.3	...	100.5	100.8	2.7	2.7	2.7	2.7	2.7	2.7	2.6	2.6
NUTS II	Norte	767.9	802	811.2	820.8	823.1	...	87.6	89.2	2.4	2.3	2.3	2.4	2.4	2.3	2.3	2.3
NUTS III	Alto Minho	786.8	817.5	813.4	816	829.9	...	73.6	77.6	2.9	3	3	3	3	2.7	2.7	2.7
Município	Arcos de Valc	734.4	757.7	764.5	789.9	802.3	...	55.8	62	3.6	3.6	3.6	3.4	3.4	3.1	3.1	3.1
Município	Caminha	879.2	911.4	854.9	846.2	843.1	...	80.3	81.8	3.2	3.1	3.1	2.8	2.8	2.6	2.6	2.6
Município	Melgaço	729.9	766.3	753.1	735.7	757.2	...	55.8	62.5	3.2	3.9	3.9	3.5	3.5	3.1	3.1	3.1
Município	Monção	718.7	769.3	768.1	773.5	786.5	...	62.9	69.4	2.7	2.7	2.7	3	3	2.9	2.9	2.9
Município	Paredes de Ci	713.7	731.1	740	733.7	751.1	...	56.1	59.2	2.2	3.7	3.7	3.6	3.6	2.8	2.8	2.8
Município	Ponte da Bari	753.8	790.1	789.5	796.3	792.4	...	53.4	60.1	3.4	3.3	3.3	3.2	3.2	2.9	2.9	2.9
Município	Ponte de Lim	747.8	763.3	751.2	758.4	773	...	62.8	65	3.2	3	3	2.9	2.9	2.8	2.8	2.8
Município	Valença	824.8	847.7	862.9	854.8	876.9	...	80.9	85.1	3.2	4	4	3.5	3.5	3.3	3.3	3.3
Município	Viana do Cast	901.2	916.7	944.1	943.7	957.3	...	89.7	93.1	2.5	2.7	2.7	2.7	2.7	2.4	2.4	2.4
Município	Vila Nova de	864.4	921.7	905.5	927.3	958.9	...	75.1	79.9	2.6	3	3	2.8	2.8	2.3	2.3	2.3
NUTS II	Centro	818.2	851.5	859.2	869.2	869	...	84.4	87.5	2.9	3.4	3.4	3.2	3.2	3.1	3.1	3.1
NUTS II	Área Metropo	1105.5	1152.9	1173.7	1195	1169.6	...	134.2	131	2.6	2.4	2.4	2.4	2.4	2.2	2.2	2.2
NUTS II	Alentejo	868.5	905.7	914.7	927.4	929.4	...	88.4	88	3.7	3.7	3.7	3.7	3.7	3.4	3.4	3.4
NUTS II	Algarve	849.1	877.1	891.8	887	880	...	100.4	96.7	2.5	2.8	2.8	2.8	2.8	2.6	2.6	2.6
NUTS I	Região Autón	954.9	982.1	1031.5	1025.1	1068.8	...	94.7	85	1.2	1.3	1.3	1.4	1.4	1.4	1.4	1.4
NUTS II	Região Autón	954.9	982.1	1031.5	1025.1	1068.8	...	94.7	85	1.2	1.3	1.3	1.4	1.4	1.4	1.4	1.4

Tabela 1 – Exemplo de dados do Dataset processado