Coimbra
Business School
Politécnico de Coimbra

Unidade Curricular de Data Mining

Mestrado em Análise de Dados e Sistemas de Apoio à Decisão

Ano Letivo 2021 - 2022

Evolução dos preços do gás natural em Portugal comparando com os países da UE

Autor(es)

Fabíola de Souza Queiroz

Maria Inês Vicente

Oleksandra Kukharska

Elaborado em

30/06/2022

Índice

Índice de figuras

Índice de tabelas

Lista de siglas e acrónimos

DCBD - Descoberta de Conhecimento em Base de Dados

CRISP-DM - Cross-Industry Standard Process for Data Mining

UE - União Europeia

DGEG - Direção Geral de Energia e Geologia

KDD - *Knowledge Discovery in Databases*

CDD - Índice do grau de arrefecimento diário (em inglês, *Cooling Degree Days*)

HDD - Índice de graus de aquecimento diário (em inglês, *Heating Degree Days*)

Introdução

Para o projeto DCBD, utilizamos dados abertos da Eurostat (Gabinete de Estatísticas da União Europeia), seguindo a metodologia de *data mining* CRISP-DM e outras técnicas para a explicação do tarifário do Setor do Gás Natural em Portugal e traçar uma previsão. O objetivo é fomentar a existência de padrões utilizando agentes de mercado (variáveis explicativas) e comparar ao cenário atual dos demais países da UE considerando a situação geopolítica existente. De acordo com a ilustração abaixo, escalamos o tema analisado e de acordo com as etapas adequamos as pesquisas no fluxo proposto.

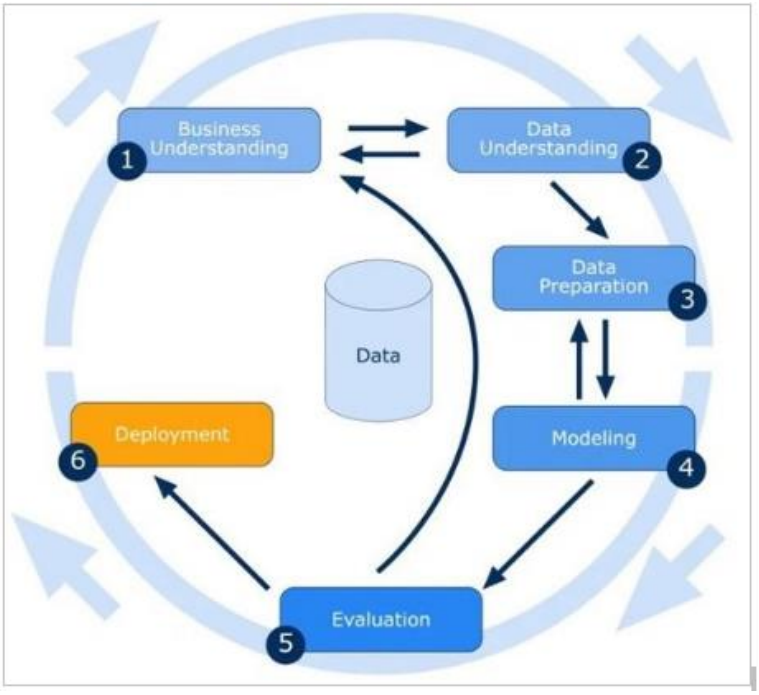

Figura 1 - Etapas da metodologia CRISP-DM

Fonte: Dataprev

Etapa 1- *Entendimento do Negócio (Business Understanding):* O problema de negócio em tese visa analisar o aumento dos preços e do consumo do gás natural, fazendo uma comparação a nível da União Europeia, levando em conta o cenário da atual situação geopolítica na Europa da Guerra Ucrânia *versus* Rússia.

Etapa 2 - *Entendimento dos Dados (Data Understanding):* Existem vários fatores que impactam diretamente nos preços do gás natural, mas, após um estudo do objetivo geral constatou-se características significativas que alteram o tarifário: Graus-dias de aquecimento e resfriamento, importação e exportação, banda de consumo e a evolução histórica dos preços.

Etapa 3 - *Preparação dos Dados (Data Preparation):* Para preparar os dados para a modelagem avaliamos todas as bases de dados abertos sobre o tema e foi possível

obter os dados brutos iniciais de uma única fonte (Eurostat), separados por variável, que passaram pela limpeza e transformação na próxima etapa;

Etapa 4 - *Modelagem (Modeling):* Aplicar técnicas de modelagem com o problema a ser resolvido (Próxima etapa do projeto).

Etapa 5 - *Avaliação (Evaluation):* Realização de testes com o modelo gerado para validar se atendem às necessidades do negócio (Próxima etapa do projeto).

Etapa 6 - *Utilização ou Aplicação (Deployment):* Apresentação dos resultados da modelagem para a tomada de decisão (Próxima etapa do projeto).

1- Entendimento do Negócio

1.1 Determinar objetivos do Negócio

A versatilidade do gás natural estimula e facilita a economia por isso a procura cada vez mais elevada a nível mundial, assim como o interesse público na regulação desses mercados extremamente elevados.

Este projeto tem o intuito de demonstrar como a regulação do preço permite avaliar a evolução do setor energético, considerando determinados cenários com dados estatísticos da Eurostat a nível da UE sobre decisões e planejamentos para avaliar o desenvolvimento em termos de energia, além destacar a harmonização das estatísticas em estreita cooperação com as autoridades nacionais.

O objetivo será, então, analisar a evolução dos preços com base em variáveis que descrevem as mudanças de comportamento de consumo, importação e exportação válidas para este problema.

1.1.1 Variável de Saída

A evolução dos preços do gás natural para consumidores domésticos e não domésticos em Portugal em relação aos demais países União Europeia (UE) considerando a situação geopolítica atual.

1.2 Funcionamento do Mercado de Gás Natural

O setor de gás natural dos países europeus, delineado por consumidores e fornecedores, deve estabelecer entidades reguladoras que administram a exploração, transporte, instalações de armazenamento, terminais e a distribuição de gás. Sabemos que Portugal não possui reservas de gás natural e importa a totalidade do que consome e possui a menor capacidade de armazenamento. Um mercado de gás padrão onde há produção, consumo e fornecimento segue o seguinte processo de regulação:

- Produção: Extração do subsolo através de poços;
- Transporte e armazenamento: Transportado em embarcações e encaminhado para pontos de armazenamento.
- Distribuição e consumo: As redes de distribuição asseguram o abastecimento direto dos consumidores domésticos e empresas, e por meio de contadores de gás permitem informar ao consumidor a energia consumida para faturação.

Figura 2 - Funcionamento do gás, desde o armazenamento até ao consumidor

Fonte: ERSE - Entidade Reguladora dos Serviços Energéticos - "GÁS: COMO FUNCIONA?"

As negociações em Portugal são feitas em sua maioria através de cláusulas *take-or-pay,* onde determinam-se contratualmente quantidades mínimas de gás natural que satisfaçam a viabilidade económica e incentivo para que não haja a entrada de contratos com custos mais elevados dos mercados internacionais, priorizando o gás natural com custos mais baixos. Assim, as existências destes tipos de contratos de longo prazo induzem à valorização, levanta questões de competitividade e consequentemente impacta na formação do preço do gás natural, e esta diretriz é adotada ao gás entregue em Portugal.
Ainda no cenário de funcionamento deste mercado, além da regulação por entidades específicas e do tipo de negociação para aquisição é importante apontar a banda de consumo, que é uma métrica que determina um limite de utilização e padroniza um tarifário para o consumidor final.

1.3 Cenário Atual Portugal e UE

Portugal, no ano de 2021, foi um dos países com mais dependência para importação de gás natural da União Europeia (UE). Segundo dados estatísticos da Direção Geral de Energia e Geologia (DGEG), atualizados em 31/05/2022, o maior volume de importação ainda é assegurado pela Nigéria, que é um dos principais fornecedores de hidrocarbonetos a Portugal e mesmo com a atual guerra na Ucrânia, a exposição ao gás russo é relativamente reduzida, embora tenha crescido nos últimos anos.

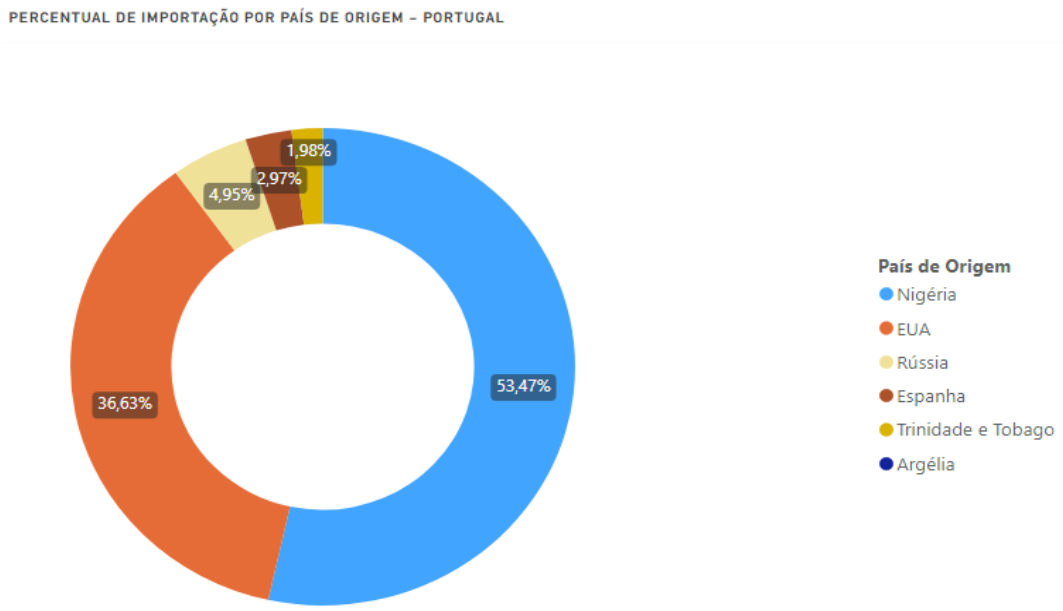

Figura 3 - Percentagem de importação de gás natural, por país de origem, em Portugal

Fonte: Elaborado pelos autores - Direção Geral de Energia e Geologia (DGEG)

Em relação ao cabaz energético que corresponde à produção de eletricidade de forma sustentável, ao olhar para o mercado de calor, verificamos que a produção de aquecimento tem aumentado regularmente e quase 100% da energia primária utilizada provém do gás natural, que não é renovável. Fazendo com que as dependências de importações representem um quarto do cabaz energético do país no ano, segundo dados da Eurostat. Apesar disso, Portugal pode ser referência em fontes de energia mais limpas e rentáveis e exercer papel importante se focar no dilema de interdependência económica da União Europeia no que diz respeito às suas necessidades energéticas.
Relativo à distribuição, são utilizadas bandas de consumo como metodologias para os mercados de gás de energia, são métricas que classificam comportamentos normais típicos numa média superior a 6 meses em vez de preços fixos e caracterizadas por faixas de consumo anual:

Para residências de gás natural*:*

- Banda-D1 (Pequena): consumo anual abaixo de 20 GJ
- Banda-D2 (Médio): consumo anual entre 20 e 200 GJ
- Banda-D3 (Grande): consumo anual acima de 200 GJ

Para indústria de gás natural:

- Banda-I1: consumo anual inferior a 1.000 GJ
- Banda-I2: consumo anual entre 1.000 e 10.000 GJ
- Banda-I3: consumo anual entre 10.000 e 100.000 GJ
- Banda-I4: consumo anual entre 100.000 e 1.000.000 GJ

- Banda-I5: consumo anual entre 1.000.000 e 4.000.000 GJ
- Banda-I6: consumo anual acima de 4.000.000 GJ (voluntário)

A mais representativa segundo estudos estatísticos, é a banda D1, a referida tarifa é reguladora de venda a clientes finais.

1.4 Definir os objetivos de Data Mining

O objetivo de Data Mining deve ser visto como um processo analítico que comporte uma grande quantidade de dados numa interação sistemática que verifica a consistência e padrões entre as variáveis escolhidas.

Dados: *Dataset* importados com extensão xlsx e csv em https://ec.europa.eu/eurostat.

Processo: Junção de várias bases de dados de diferentes variáveis explicativas.

Software para artigos: Mendeley

Linguagem de programação: Python

1.5 Produzir plano do projeto

Através dos conceitos de KDD (Knowledge Discovery in Databases) ou Descoberta de Conhecimento em Bases de Dados pôde-se tornar a análise dos dados compreensível e potencialmente útil, com suas etapas sendo:

1. Preparação dos Dados
2. Definição do Conjunto de Dados
3. Limpeza e Pré-processamento de Dados
4. Redução dos Dados
5. Mineração de Dados
6. Interpretação dos Resultados

2 - Estudo dos dados

2.1 Recolha dos dados iniciais

A recolha dos dados iniciais foi feita através do “Eurostat”, o Serviço de Estatística da União Europeia responsável pela publicação de estatísticas e indicadores de elevada qualidade a nível europeu. Com a opção “Database” é apresentado um conjunto de tabelas disponíveis relativas ao tema “Energy”.

Como o interesse inicial se prende com os preços do gás natural no consumo doméstico da banda D1, com todas as taxas e impostos associados, é necessário selecionar todas as bases de dados relativas ao tema.

Neste projeto optámos por selecionar, um total de seis bases de dados, nomeadamente:

- **“Gas prices for household consumers - bi-annual data (from 2007 onwards)”**: os preços do gás natural de consumo doméstico e não doméstico;
- **“Supply, transformation and consumption of gas - monthly data”**, mais concretamente a **“Energy balance: inlance consumption”**: o consumo interno de gás natural por país;
- **“Imports of natural gas by partner country - monthly data”**: as importações de gás natural por país;
- **“Exports of natural gas by partner country - monthly data”**: as exportações de gás natural por país**;**
- **“Cooling degree days by country - monthly data”**: o índice de grau de arrefecimento diário por país;
- **“Heating degree days by country - monthly data”**: o índice de grau de aquecimento diário por país.

2.2 Descrição dos dados

2.2.1 Preços do gás natural para consumo doméstico - dados bi-anuais (a partir de 2007)

Este *dataset* contém dados semestrais relativos aos preços do gás natural no consumo doméstico em Euro/KWH, desde o segundo semestre de 2007 ao segundo semestre de 2021, com todas as taxas incluídas, na banda de consumo D1 com o consumo anual menor que 20GJ, a mais representativa na União Europeia, inclusive Portugal, segundo o entendimento do negócio.

2.2.2 Oferta, transformação e consumo de gás natural - dados mensais

2.2.2.1 Consumo interno de gás natural

O *dataset* contém dados mensais do consumo interno de gás natural na União Europeia e em Portugal, desde janeiro de 2014 a março de 2022, em milhões de metros cúbicos.

2.2.2.2 Importações de gás natural

São disponibilizados dados mensais das importações de gás natural na União Europeia e em Portugal, desde janeiro de 2014 a março de 2022, em milhões de metros cúbicos.

2.2.2.3 Exportações de gás natural

Tal como o consumo interno e as importações, neste dataset são disponibilizados dados mensais das exportações de gás natural na União Europeia e em Portugal, desde janeiro de 2014 a março de 2022, em milhões de metros cúbicos.

2.2.3 Índices de graus de arrefecimento e aquecimento diários por país - dados mensais

2.2.3.1 Índice de grau de arrefecimento diário

Este *dataset* é composto por um índice do grau de arrefecimento diário (CDD), que é um índice técnico baseado no tempo concebido para descrever a necessidade dos requisitos de arrefecimento dos edifícios, que quantifica a procura por ar condicionado.

Os dias com graus de aquecimento são calculados durante um período de tempo (normalmente um ano), através da soma das diferenças entre a temperatura média diária de cada dia e a temperatura de 18°C. Para qualquer temperatura acima de 18°C, assume-se que o edifício não necessita de qualquer aquecimento. Por exemplo, três dias de Inverno consecutivos com temperaturas médias de 4°C, -2°C e -4°C totalizam 56 HDD.

É composto por dados mensais de janeiro de 1979 a dezembro de 2021 e os dados CDD são apresentados como somas de temperaturas em ˚C.

2.2.3.2 Índice de grau de aquecimento diário

O índice de graus-dia de aquecimento (HDD) é um índice técnico baseado no tempo, concebido para descrever a necessidade das condições de energia de aquecimento dos edifícios, derivado de medições de temperatura de ar externo. Desta forma, um exemplo deste índice é o facto de quando obtemos três dias de Verão de 26°C, 28°C, e 30°C, estes totalizam 30 CDD. Tal como o CDD, é composto por dados mensais de janeiro de 1979 a dezembro de 2021 e também são apresentados como somas de temperaturas em ºC.

2.3 Exploração dos dados

De seguida é apresentado um conjunto de estatísticas sumárias, através da linguagem de programação em Python, que permitem aprofundar as informações das variáveis disponibilizadas nos *dataset*.

2.3.1 Análise Univariada

2.3.1.1 Preços do gás natural para consumo doméstico da banda de consumo D1

De acordo com as estatísticas principais, existem 29 instâncias para Portugal e 30 instâncias para a União Europeia em termos de preços do gás natural para consumo doméstico da banda de consumo D1. A média em ambos está por volta dos 0.097 Euro/KWH, com um desvio padrão de 0.014 e 0.011 para Portugal e União Europeia, respetivamente.

Portugal

```
count    29.000000
mean      0.096279
std       0.013790
min       0.073800
25%       0.083300
50%       0.095400
75%       0.105600
max       0.125600
Name: OBS_VALUE, dtype: float64
```

União Europeia

```
count    30.000000
mean      0.097830
std       0.011063
min       0.079600
25%       0.089400
50%       0.097050
75%       0.107425
max       0.114500
Name: OBS_VALUE, dtype: float64
```

Figura 4 - Principais estatísticas sumárias dos preços do gás natural para consumo doméstico da banda de consumo D1, para Portugal e União Europeia

2.3.1.2 Consumo interno de gás natural

O consumo interno de gás natural contém 99 instâncias para Portugal e 98 para a União Europeia, com uma média mais acentuada na União Europeia devido à concentração de consumo nos 27 países, com desvio padrão maior, o que significa haver uma maior dispersão em relação aos diferentes países.

Portugal

```
count     99.000000
mean     454.495616
std       81.417860
min      281.000000
25%      386.046000
50%      462.000000
75%      519.395500
max      600.000000
Name: OBS_VALUE, dtype: float64
```

União Europeia

```
count       98.000000
mean     32543.225745
std      10727.598285
min      18431.000000
25%      22413.500000
50%      30280.293000
75%      42053.250000
max      57611.000000
Name: OBS_VALUE, dtype: float64
```

Figura 5 - Consumo interno de gás natural, para Portugal e União Europeia

2.3.1.3 Importações de gás natural

A importação de gás natural contém 99 instâncias para Portugal e 98 para a União Europeia. Portugal é um país que, atualmente, importa maioritariamente o seu consumo de gás natural, portanto apresenta-se com uma média de 474 milhões de metros cúbicos de importações de gás natural. O desvio padrão é maior na União Europeia devido às diferenças de importações entre países. Em Portugal, é também notável essa diferença, sendo relativamente menor visto se tratar apenas de um país.

Portugal

```
count     99.000000
mean     474.794515
std       99.193817
min      226.000000
25%      401.000000
50%      471.000000
75%      547.000000
max      696.000000
Name: OBS_VALUE, dtype: float64
```

União Europeia

```
count       98.000000
mean     52584.100327
std       4529.644616
min      41279.000000
25%      49143.145000
50%      52249.837500
75%      55467.667500
max      63125.410000
Name: OBS_VALUE, dtype: float64
```

Figura 6 - Importações de gás natural, para Portugal e para a União Europeia

2.3.1.4 Exportações de gás natural

O valor das exportações de gás natural é escasso para Portugal, com uma média de apenas 17 milhões de metros cúbicos, visto este ser um país que importa maioritariamente as quantidades de gás natural. Comparativamente, a União Europeia apresenta uma média de 27044 milhões de metros cúbicos em exportações de gás natural, com um desvio padrão bastante acentuado.

Portugal

```
count     96.000000
mean      17.295844
std       27.376359
min        0.000000
25%        0.000000
50%        3.000000
75%       32.250000
max      104.000000
Name: OBS_VALUE, dtype: float64
```

União Europeia

```
count       98.000000
mean     27044.010235
std       2998.717890
min      20778.655000
25%      25296.472250
50%      26636.000000
75%      28412.750000
max      37434.000000
Name: OBS_VALUE, dtype: float64
```

Figura 7 - Exportações de gás natural, para Portugal e União Europeia

2.3.1.5 Índice de grau de arrefecimento diário

Através das médias, podemos observar que, para um determinado edifício, houve uma maior necessidade de utilização de ar condicionado em Portugal comparativamente à União Europeia, devido às altas temperaturas ao longo dos anos.

Portugal

```
count    516.000000
mean      15.131822
std       26.395734
min        0.000000
25%        0.000000
50%        0.000000
75%       20.352500
max      128.980000
Name: OBS_VALUE, dtype: float64
```

União Europeia

```
count    516.000000
mean       6.089205
std       11.291459
min        0.000000
25%        0.000000
50%        0.045000
75%        6.607500
max       60.270000
Name: OBS_VALUE, dtype: float64
```

Figura 8 - CDD para Portugal e União Europeia

2.3.1.6 Índice de grau de aquecimento diário

Em Portugal, a necessidade de aquecimento em edifícios foi menor comparativamente aos resultados da União Europeia.

Portugal

```
count    516.000000
mean     103.250756
std       99.647194
min        0.000000
25%        4.710000
50%       74.170000
75%      185.912500
max      330.780000
Name: OBS_VALUE, dtype: float64
```

União Europeia

```
count    516.000000
mean     267.230233
std      187.948832
min        8.180000
25%       79.900000
50%      257.310000
75%      433.940000
max      720.410000
Name: OBS_VALUE, dtype: float64
```

Figura 9 - HDD para Portugal e União Europeia

2.4 Verificação da qualidade dos dados

Em relação ao *dataset* “Gas prices for household consumers”, em 2008, foi elaborado um relatório de autoavaliação sobre o processo de recolha de preços de gás e eletricidade. Como a metodologia para a recolha de dados sobre os preços do gás está bem definida, o conjunto de dados fornece informações sobre preços comparáveis entre países. Contudo, os utilizadores devem estar cientes de que sempre que os preços do gás em euros para países não europeus são comparados no tempo, os efeitos das diferenças nas taxas de câmbio também são tidos em conta. Sempre que os preços são calculados e/ou apresentados em euros, às taxas de câmbio médias dos dois trimestres do semestre apropriado são tomadas como referência.

De acordo com os *dataset* “Inlance consumption”, “Imports” e “Exports” o Eurostat realiza testes de qualidade, principalmente sobre a coerência da informação fornecida. Além disso, os questionários utilizados para a transmissão de dados também são incorporados em testes de coerência. Este conjunto de dados faz parte

das estatísticas energéticas definidas no Regulamento (CE) n.º 1099/2008 sobre estatísticas energéticas. É considerado como estatísticas europeias e, consequentemente, aplica-se o quadro do SEE para a qualidade. Além disso, estão integrados no ciclo de Relatórios de Qualidade que tem lugar de cinco em cinco anos. Os relatórios de qualidade para as estatísticas da energia baseiam-se no artigo 6º do Regulamento (CE) nº 1099/2008 relativo às estatísticas da energia.

Finalmente, para os *dataset* “Cooling and heating degree days by country”, este conjunto de dados inclui os dados mensais publicados pelo Portal de Recursos AGRI4CAST do Centro Comum de Investigação. O Eurostat não é o produtor dos dados mensais, está apenas a re-publicá-los.

Aquando da exploração dos dados, estes estão guardados em excel e csv, de forma a proceder ao tratamento destes com a programação em Python. Foram detetados dados em falta em todos os *dataset*, pelo que teremos de proceder à preparação dos dados, nomeadamente à limpeza, construção e integração.

3 - Preparação dos Dados

Nesta etapa, os dados relativos às seis variáveis são preparados para os algoritmos de modelagem. Apesar dos dados já estarem no formato tabular, onde cada instância representa o período temporal e o país/organização, foi preciso selecionar um período de tempo e unidade de análise, tratar dos dados em falta, verificar se existem ou não *outliers* que eventualmente podem prejudicar os resultados de análise. E também fazer alterações quanto à estruturação, como por exemplo a fusão dos dados.

3.1 Seleção dos Dados

A seleção dos dados refere-se à seleção de colunas e linhas numa tabela. Não houve necessariamente necessidade de excluir colunas, no entanto, os dados relativos à União Europeia e a Portugal irão ser os mais utilizados para a análise.

Na análise de dados serão utilizadas as seguintes variáveis:

- **price**: “Gas prices for household consumers - bi-annual data (from 2007 onwards)”;
- **inlance**: “Energy balance: inlance consumption”;
- **imports**: “Imports of natural gas by partner country - monthly data”;
- **exports**: “Exports of natural gas by partner country - monthly data”;
- **cooling**: “Cooling degree days by country - monthly data”;
- **heating**: “Heating degree days by country - monthly data”.

Das seis variáveis selecionadas para análise, apenas a variável **price** tem como unidade o semestre, ao contrário das restantes cinco variáveis, que têm unidade mensal, tendo-se neste caso, optado pela realização da conversão dos dados mensais para semestrais.

Não foi necessário realizar testes de significância para decidir que campos incluir, nem recorrer a dados adicionais, uma vez que, todos os *dataset* contém a informação necessária à elaboração da análise.

3.2 Limpeza dos Dados

A limpeza dos dados é um passo importante na preparação dos dados visto que nos ajuda a garantir a qualidade dos dados.

3.2.1. Período Temporal

Os *dataset* selecionados para além de não se encontravam na mesma unidade, também não se encontravam no mesmo período de tempo, como se observa na tabela seguinte. Só foram selecionados os dados a partir de 2014-01 até 2021-12, tendo os dados sido posteriormente convertidos para a unidade semestral.

Variáveis	Período Temporal	Unidade
price	2007 S1 a 2021 S2	Semestral
inlance	2014-01 a 2022-04	Mensal
imports	2014-01 a 2022-04	Mensal
exports	2014-01 a 2022-04	Mensal
cooling	1979-01 a 2021-12	Mensal
heating	1979-01 a 2021-12	Mensal

Tabela 1 - Período Temporal e Unidade das Variáveis

3.2.2. Dados em falta

É necessário representar os dados em falta de alguma maneira para posteriormente conseguir executar algum modelo. Uma das alternativas é a exclusão da observação/coluna, porém, este processo irá reduzir significativamente o tamanho da base de dados, comprometendo a sua viabilidade, deste modo, optou-se por substituir estes valores por zero (0) em quatro dos seis *dataset*, pois, os dados dizem respeito ao valor monetário de cada variável.

A tabela seguinte, mostra a quantidade de valores NaN existentes nos *dataset*:

Variáveis	Nº de dados NaN
price	58
inland	542
imports	565
exports	696
cooling	0
heating	0

Tabela 2 - Dados em Falta nas Variáveis

3.2.3 Deteção de Outliers

Nos dados relativos a Portugal, as variáveis **cooling** e **price** apresentam ambas *outliers*, em que os quartis apresentam uma diferença entre o terceiro quartil para o valor máximo.

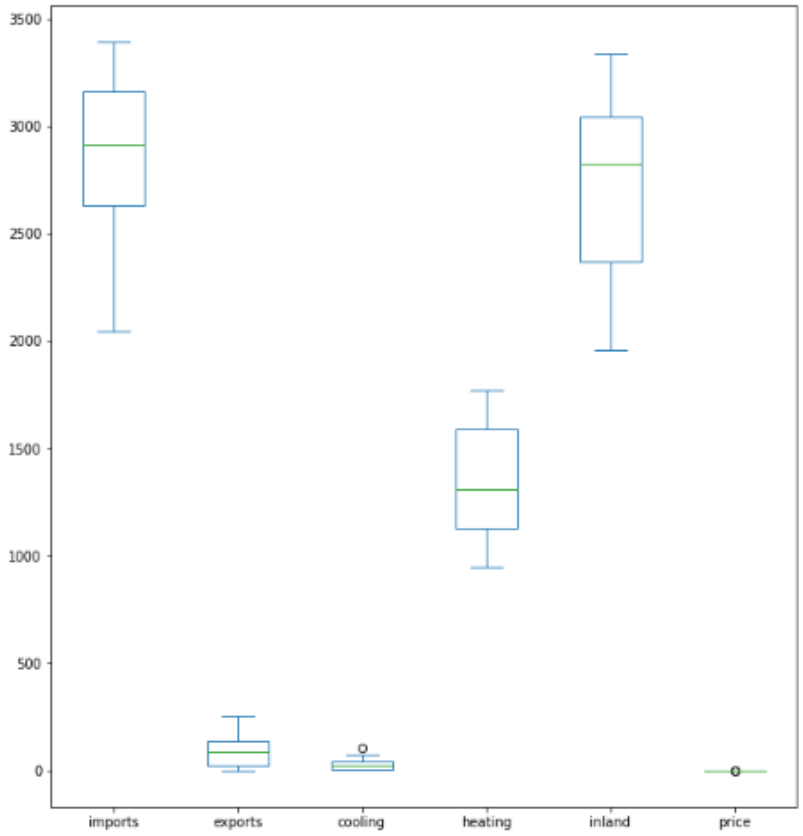

Figura 10 - Box-Plot da identificação de outliers em Portugal

cooling apresenta um outliers

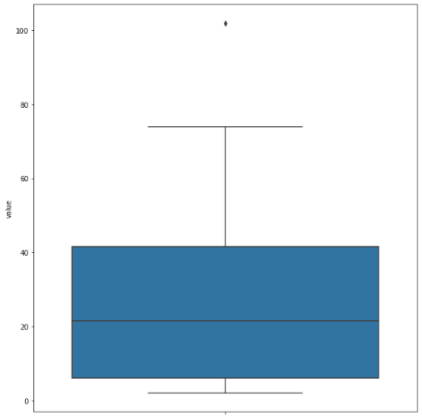

Figura 11 - Box-Plot referente à variável cooling, em Portugal

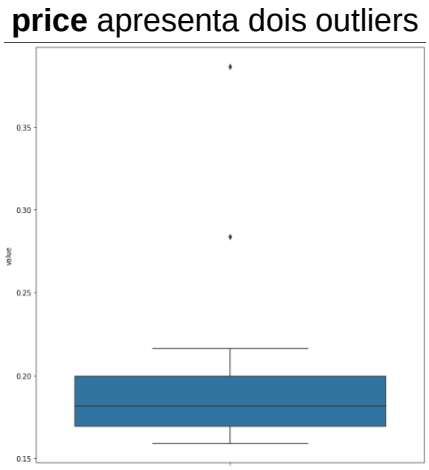

Figura 12 - Box-Plot referente à variável price, em Portugal

Consoante o objetivo da análise a ser feita, a existência de *outliers* pode provocar o enviesamento do resultado. Neste caso, optou-se por ignorá-los, não os excluir da amostra de dados, uma vez que os dados são semestrais, o que iria afetar significativamente o tamanho da base de dados. No entanto, foi preferível tê-los em consideração, caso seja necessário realizar uma análise separada ou utilizar métodos de clusterização.

3.3 Construção e arquitetura dos Dados

Depois de harmonizar a informação constante nos diferentes *dataset* tendo em vista a utilização de uma única unidade de análise (semestre) e após o tratamento dos dados em falta, estamos em condições de proceder agrupamento das variáveis.

Uma vez que, o propósito é fazer a comparação do preço do gás entre Portugal e os países da UE, de forma a facilitar o acesso à informação, os valores referentes às seis variáveis serão agrupados numa única tabela.

O *output* a seguir apresentado, mostra a *head()* da tabela relativa aos campos e valores de Portugal que serão utilizados no modelo:

Portugal

	imports	exports	cooling	heating	inland	price
2014-S1	2186.0	253.0	5.73	1402.15	1955.0	0.1671
2014-S2	2267.0	90.0	6.05	943.57	2155.0	0.1739
2015-S1	2433.0	248.0	4.14	1600.07	2324.0	0.1703
2015-S2	2694.0	159.0	101.91	1102.99	2387.0	0.1874
2016-S1	2045.0	0.0	8.48	1581.23	2149.0	0.1651

Figura 13 - Agrupamento de variáveis

3.4 Integração e relacionamento dos Dados

Visto que cada *dataset* refere-se a uma variável, que contém os valores para a União Europeia e os países que a integram, será feita uma agregação dos dados, que consiste em criar um novo registo que resulta da informação sumária de um conjunto de registos anterior de uma tabela, de forma a simplificar os dados, serão agregados os dados da União Europeia e de Portugal numa única tabela, como se apresenta na tabela seguinte.

	União Europeia	Portugal
2014 S1	291481.0	2186.0
2014 S2	277064.0	2267.0
2015 S1	279607.0	2433.0
2015 S2	300655.0	2694.0
2016 S1	297527.0	2045.0

Figura 14 - Agregação dos dados

3.5 Formatação dos Dados

Nesta fase pretende-se explicar todas as alterações por nós efetuadas, aos dados, de modo a clarificar e torná-las mais entendíveis. No entanto, não foi necessário fazer nenhuma alteração sintática dos dados.

4 - Modelação

4.1 Seleção de Técnica de Modelação

4.2 Geração de desenhos e teste

4.3 Construção do Modelo

4.4 Revisão do modelo

5 - Avaliação

5.1 Avaliação de resultados

5.2 Revisão do processo

5.3 Determinação dos passos seguintes

Conclusão

Referências

http://desenvolvimento.dataprev.gov.br/pddataprev_internet/visualizar_guia.php?idguia=700