



ETL Rede Elétrica Barcelos

Relatório Trabalho Prático

Integração de Sistemas de Informação

27977 – Igor Miguel Torres da Costa

Professor: Óscar Rafael da Silva Ferreira Ribeiro

Licenciatura em Engenharia de Sistemas Informáticos

Barcelos | outubro, 2025

Lista de Abreviaturas e de Siglas

- **ETL:** *Extract, Transform, Load*
- **UPAC:** Unidades de Produção e Autoconsumo
- **API:** *Application Programming Interface*
- **JSON:** *JavaScript Object Notation*
- **HTTP:** *Hypertext Transfer Protocol*

Índice

1. Problema	7
2. Estratégia utilizada	8
2.1. Extração (<i>Extract</i>)	8
2.2. Transformação (<i>Transform</i>)	11
2.3. Carregar (<i>Load</i>)	12
3. Transformações.....	13
4. Jobs.....	22
5. Vídeo com demonstração.....	24
6. Conclusão	25
7. Bibliografia.....	26

Índice de Figuras

Figura 1 - Configuração do Node Excel Reader	9
Figura 2 - Parte Componente Consumo Mensal Get Request	10
Figura 3 - Configuração GET Request Node	10
Figura 4 - Variável Flows	10
Figura 5 - Componente JSON Path	11
Figura 6 - Componentes String Manipulation, Ungroup e Column Filter	11
Figura 7 - Componente Joiner	12
Figura 8 - Componente GroupBy	12
Figura 9 - Componente Math Formula	12
Figura 10 - Componente Visualização dos Dados	13
Figura 11 - Componentes de Visualização e Gráficos.....	13
Figura 12 - Configuração do Componente JSON Path.....	14
Figura 13 - Configuração do Componente Ungroup.....	14
Figura 14 - Configuração do componente Column Filter	15
Figura 15 - Diagrama Transformação dos dados JSON para Tabela KNIME.....	15
Figura 16 - Tabela de Dados Pontos de Carregamento por Freguesia	16
Figura 17 - Configuração Manual do Componente GroupBy potencia_maxima_adminissiveis.....	16
Figura 18 - Configuração Manual GroupBy num_instalacao	17
Figura 19 - Configuração do Joiner para os Pontos de Carregamento.....	18
Figura 20 - Configuração Math Formula para Média de Potência por Num Instalação.....	19
Figura 21 - Tabela Transformada de Dados Postos Carregamento	19
Figura 22 - Componente Top K Row Filter.....	20
Figura 23 - Configuração do Top K Row Filter	20
Figura 24 - Diagrama Visualização de Dados	21
Figura 25 - Gráfico de Barras Média de Potência por Ligação	22
Figura 26 - Diagrama Extração de Dados com Loop	22

Figura 27 - Excerto Dashboard Interativo.....	23
---	----

1. Problema

Neste trabalho prático para a cadeira de Integração de Sistemas de Informações, pretende-se experimentar e usar as ferramentas para criar processos ETL (*Extract, Transform, Load*).

Para a realização do trabalho foi definido um problema-base: As câmaras municipais de hoje necessitam cada vez mais de agregar dados de diferentes fontes num único sistema ou plataforma, de forma a obter informação atualizada e de fácil acesso sobre as diversas áreas de competência do município. Esta integração permite melhorar o conhecimento sobre o estado dos serviços públicos e para obter informações para apoiar tomada de decisões.

Neste contexto foi escolhido o setor da rede elétrica de Barcelos, procurando demonstrar como uma câmara municipal pode obter e analisar dados relevantes sobre o consumo energético, produção, iluminação pública e mobilidade elétrica.

O objetivo é desenvolver um processo ETL que integre diferentes fontes de dados públicas e produza tabelas consolidadas com informações úteis sobre:

- Caracterização da Iluminação Pública;
- Consumo de energia elétrica mensal por freguesia;
- Ligações á rede associadas à mobilidade elétrica;
- Pontos de entrega de Energia;
- Unidades de Produção e Autoconsumo (UPACs);

A partir desses dados, será possível criar indicadores e visualizações que auxiliem a câmara municipal na gestão da rede elétrica local.

2. Estratégia utilizada

Para o desenvolvimento do processo ETL, foi escolhida a plataforma *KNIME Analytics*, tendo o fluxo sido dividido em três fases: Extração, Transformação e Carregamento.

2.1. Extração (*Extract*)

Na fase de extração, o primeiro passo foi identificar e obter fontes de dados de interesse para o problema. A E-REDES entidade responsável pela rede de distribuição elétrica de baixa e média tensão em Portugal disponibiliza um portal de dados abertos sobre a rede elétrica, esta plataforma fornece uma API pública que permite realizar pedidos do tipo GET, recebendo em formato JSON.

Foi possível aplicar filtros por concelho, de forma a recolher apenas os dados referentes ao município de Barcelos. Para esta fase de extração foram utilizados os seguintes datasets, acedidos através de endpoints GET:

- cadastro_iluminacao_publica
- 3-consumos-faturados-por-municipio-ultimos-10-anos
- 8-unidades-de-producao-para-autoconsumo
- 9-plr-mobilidade-eletrica
- 20-caracterizacao-pes-contrato-ativo

Um exemplo de um endereço HTTP para realizar uma chamada GET:

<https://e-redes.opendatasoft.com/api/explore/v2.1/catalog/...>

São usados parâmetros para fazer a chamada API como:

- *Select*: Define quais os campos ou atributos a incluir na resposta JSON
- *Where*: Estabelece condições de filtragem (ex concelho = ‘Barcelos’, nível de tensão, tipo de tecnologia, etc.)
- *Limit*: Define o número máximo de objetos devolvidos por pedido (a E-REDES impõe um limite de 100 registos por chamada.)

A API fornecida pela E-REDES contém uma limitação que apenas só se pode obter 100 registos por chamada e já que certos pedidos é necessário mais de 100 registos para obter todo contexto de informações do concelho de Barcelos.

Para ultrapassar esta limitação, foi adotada uma estratégia alternativa de extração. Foi obtido um ficheiro Excel contendo a lista completa das freguesias de Portugal, do qual foram extraídos apenas os nomes das freguesias pertencentes ao concelho de Barcelos como podemos ver na figura 1.

Com base nessa lista, foi desenvolvido no KNIME um *loop* automatizado que executa chamadas sucessivas à API para cada freguesia de Barcelos, em cada um dos endpoints selecionados. Este processo permitiu recolher todos os dados de forma sistemática, garantindo a cobertura total do concelho sem ultrapassar o limite imposto pela API, podemos ver a parte deste loop na figura 2.

E a configuração do node GET Request do KNIME na figura 3, onde podemos ver que o URL está numa variável URL *Flow* na figura 4 e que foi configurado um atraso de 500ms para evitar *timeouts* da API.

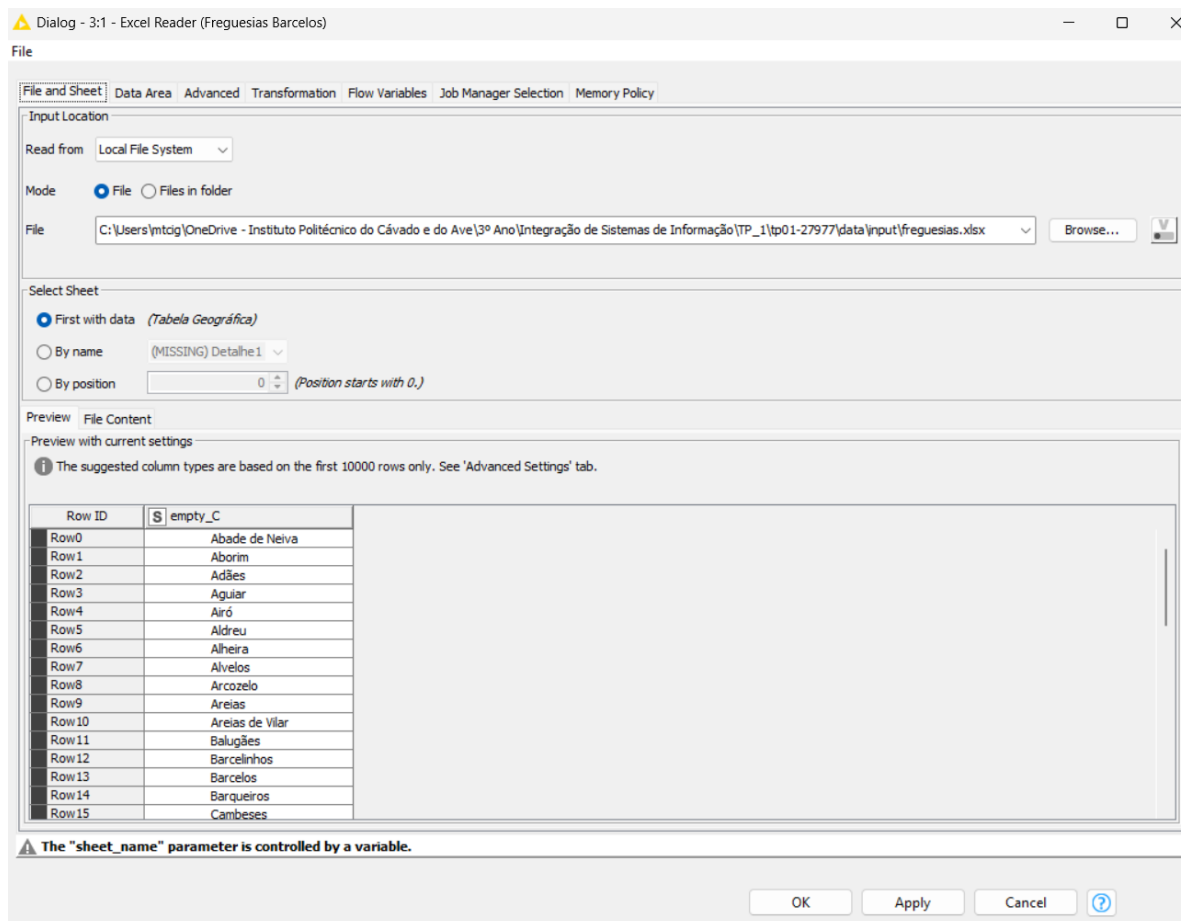


Figura 1 - Configuração do Node Excel Reader

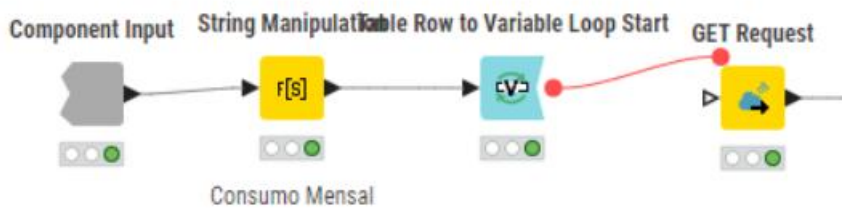


Figura 2 - Parte Componente Consumo Mensal Get Request

Dialog - 3:74:0:10 - GET Request

File

Connection | Authentication | Proxy | Error Handling | Request Headers | Response Headers | Flow Variables | Job Manager Selection | Memory Policy

☒ URL:

☐ URL column:

Handling of invalid URLs

☒ Insert missing value ☐ Fail node execution immediately

☒ Delay (ms):

Concurrency:

SSL

☐ Ignore hostname mismatches

☐ Trust all certificates

☒ Follow redirects

☒ Send large data in chunks

Connect timeout (s):

Read timeout (s):

Body column:

The "Constant URI" parameter is controlled by a variable.

Figura 3 - Configuração GET Request Node

Owner ID	Data Type	Variable Name	Value
3:74:0:8	StringType	url	https://e-redes.opendatasoft.com/api/explore/v2.1/catalog/datasets/3-consumos-faturados-por-municipio-ultimos-10-anos
3:74:0:8	StringType	RowID	Row88
3:74:0:8	IntType	currentiteration	88
3:74:0:8	IntType	maxiterations	89

Figura 4 - Variável Flows

2.2. Transformação (*Transform*)

Após a extração dos dados, foi necessário realizar diversas operações de transformação com o objetivo de limpar, uniformizar e preparar os dados para posterior análise e visualização. Devido à diversidade das fontes e à estrutura dos ficheiros JSON devolvidos pela API, alguns conjuntos de dados apresentavam campos aninhados e nomes inconsistentes, exigindo normalização.

As principais transformações realizadas incluíram:

- **Conversão de formato:** os ficheiros obtidos em JSON foram convertidos em tabelas KNIME utilizando o nó “**JSON Path**”, seguido de operações para expandir e normalizar colunas.



Figura 5 - Componente JSON Path

- **Limpeza e normalização de campos:** os nomes das freguesias e dos concelhos foram padronizados com o nó “**String Manipulation**”, removendo diferenças de acentuação e capitalização. E para a limpeza das tabelas foi usado o “**Ungroup**” desagrupar os dados JSON e dividir cada objetivo como uma linha na tabela. Já o “**Column filter**” foi usado para remover as colunas desnecessárias que seria o *header* do JSON.

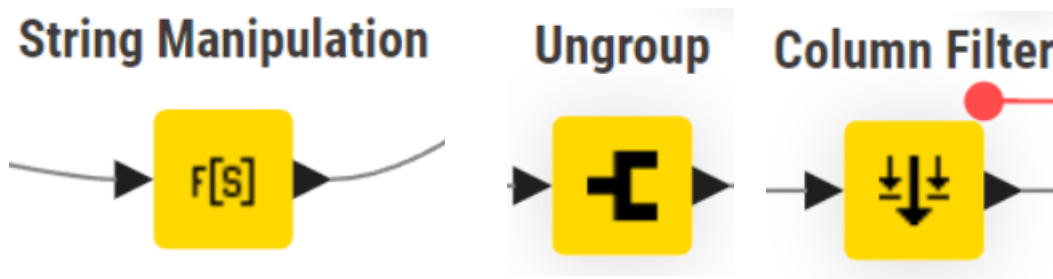


Figura 6 - Componentes String Manipulation, Ungroup e Column Filter

- **Integração de datasets:** através dos nós “**Joiner**”, foram cruzadas as várias tabelas obtidas (por exemplo, consumo energético, autoconsumo e iluminação pública), criando uma visão integrada da rede elétrica do concelho.

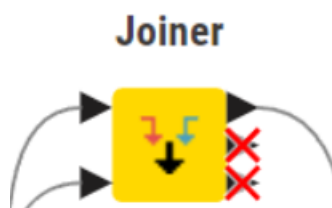


Figura 7 - Componente Joiner

- **Agregações estatísticas:** com o nó “**GroupBy**”, foram calculados indicadores relevantes, como o total de potência instalada por freguesia, o número total de luminárias por tipo de lâmpada, e o número de instalações de autoconsumo por tecnologia.

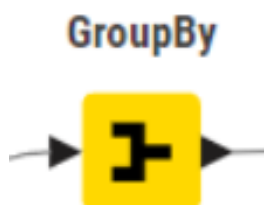


Figura 8 - Componente GroupBy

- **Criação de novos atributos:** o nó “**Math Formula**” foi utilizado para derivar métricas adicionais, como médias de consumo por ponto de entrega ou proporções entre tipos de tecnologia.



Figura 9 - Componente Math Formula

2.3. Carregar (*Load*)

Após a transformação dos dados e organizados em tabelas com dados estatísticos já calculados como soma, média, diferença. Estes dados junta-se para um componente chamado “**Visualização dos Dados**” como podemos ver na figura 10.

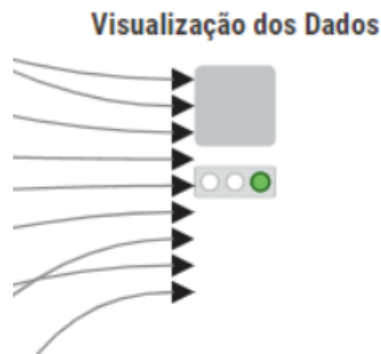


Figura 10 - Componente Visualização dos Dados

Com os dados transformados são passados para componentes de visualização de estatística como “**Bar Chart**”, “**Pie Chart**” e “**Table View**”. Ao criar um componente que junta vários componentes como o da Figura 10, podemos criar um *layout* para juntar todos os gráficos gerados numa só página, criando assim um *dashboard* interativo.

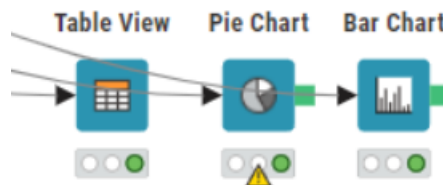


Figura 11 - Componentes de Visualização e Gráficos

3. Transformações

A primeira transformação do workflow do KNIME inicia-se após a realização da chamada GET, que devolve os dados em formato JSON. O ficheiro resultante é então processado pelo componente “JSON Path”, responsável por converter a estrutura JSON num formato tabela do KNIME selecionado apenas os dados necessários, podemos ver a configuração na figura 12.

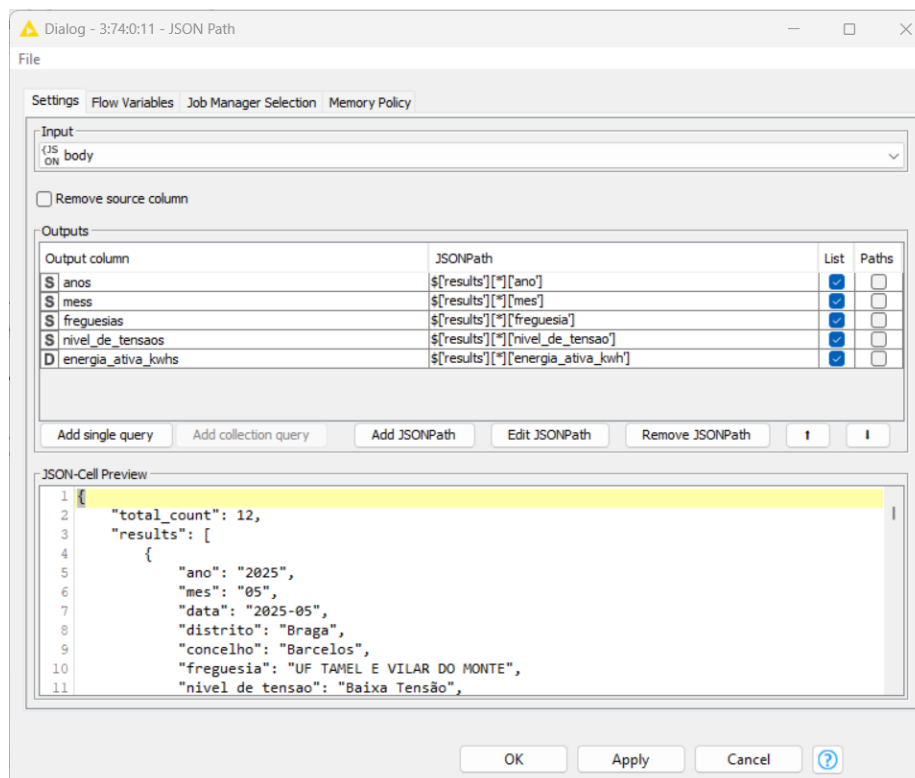


Figura 12 - Configuração do Componente JSON Path

Contudo, a conversão inicial gera uma tabela onde todos os objetos JSON se encontram agregados numa única linha, representando o conjunto completo de dados. Para resolver este problema, foi utilizado o componente “Ungroup”, que desagrega os elementos do JSON, fazendo com que cada objeto JSON corresponda a uma linha independente na tabela do KNIME, podemos ver a configuração na figura 13.

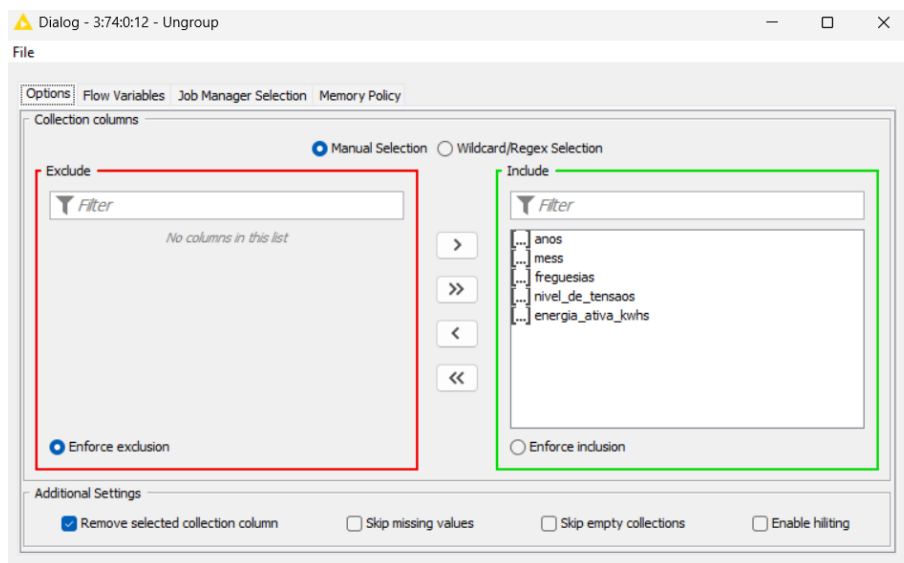


Figura 13 - Configuração do Componente Ungroup

Por fim, a tabela é submetida ao componente “Column Filter”, que remove as colunas desnecessárias, como os headers e metadados do JSON, mantendo apenas os campos relevantes para a análise, podemos ver a configuração na figura 14.

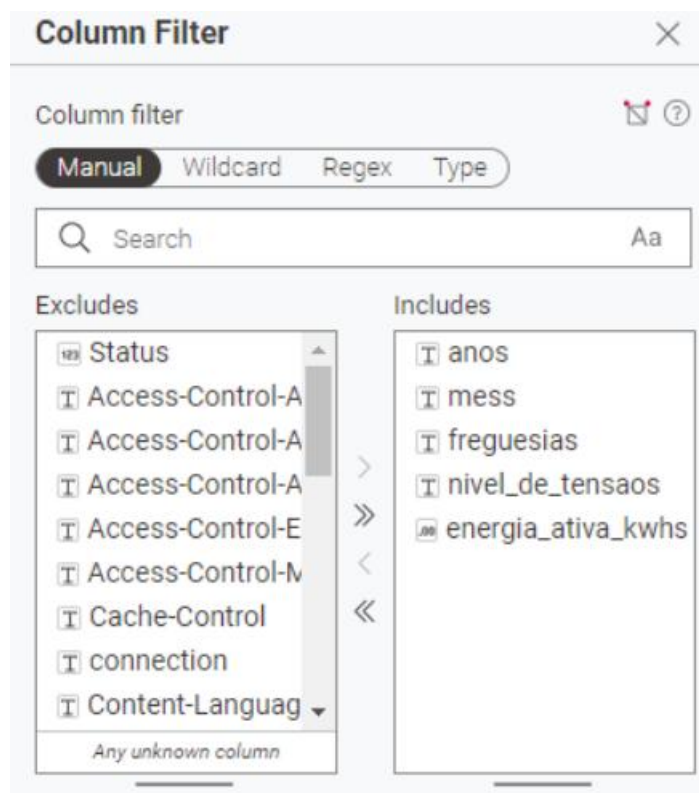


Figura 14 - Configuração do componente Column Filter

Podemos ver todo este processo no diagrama da figura 15.



Figura 15 - Diagrama Transformação dos dados JSON para Tabela KNIME

Após a transformação inicial dos dados JSON em tabelas KNIME devidamente estruturadas, procedeu-se à conversão dos dados em informações e estatísticas úteis para análise e visualização.

Como exemplo prático, pode-se considerar o conjunto de dados relativo aos Pontos de Carregamento para Mobilidade Elétrica. A partir da tabela resultante no KNIME, foram identificados os campos de maior relevância analítica, nomeadamente: **freguesias**, **potencia_maxima_admissiveis** e **num_instalacoes** como podemos ver na figura 16.

Rows: 31 | Columns: 3

#	RowID	freguesias	potencia_maxima_admissiveis	num_instalacoess
1	Row0_	Arcozelo	10.35	1
2	Row0_	Barcelos, V. Boa, V. Frescalinha	250	1
3	Row0_	Barcelinhos	90	1
4	Row0_	Arcozelo	800	1
5	Row0_	Barcelos, V. Boa, V. Frescalinha	1,850	1
6	Row0_	Barcelos, V. Boa, V. Frescalinha	69	1
7	Row0_	Barcelinhos	160	1
8	Row0_	Arcozelo	70	1
9	Row0_	Gamil e Midões	200	1
10	Row0_	Alheira e Igreja Nova	945	2
11	Row0_	Perelhal	20.7	1

Figura 16 - Tabela de Dados Pontos de Carregamento por Freguesia

Para obter informações de interesse a partir da tabela da figura 16, foi utilizado o componente “GroupBy”, Neste caso, os dados foram agrupados por freguesia, permitindo calcular a soma da potência máxima admissível dos postos de carregamento, como podemos ver na figura 17.

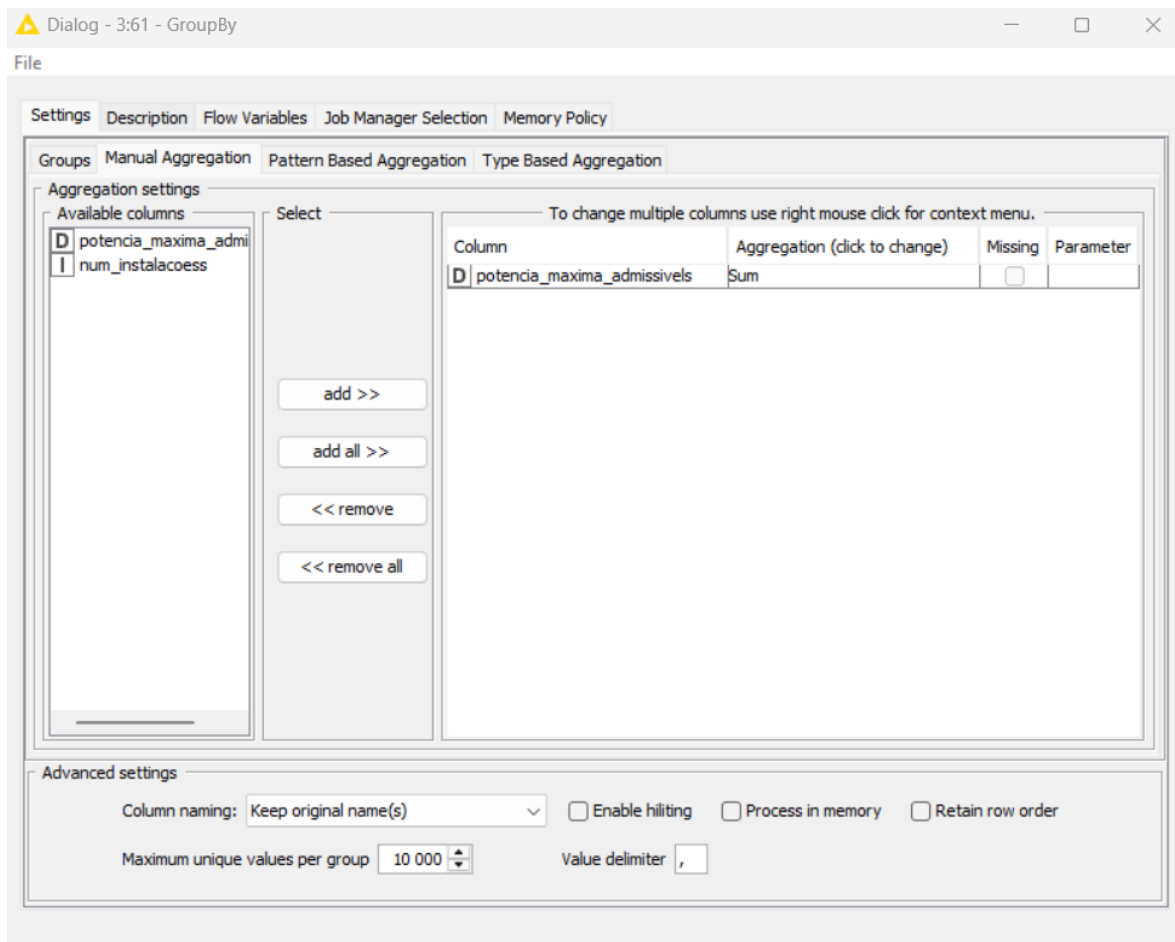


Figura 17 - Configuração Manual do Componente GroupBy potencia_maxima_admissiveis

Já para o segundo componente “GroupBy”, aos dados são agrupados por freguesia, que permite calcular o número de instalações de postos de carregamento, como podemos ver na figura 18.

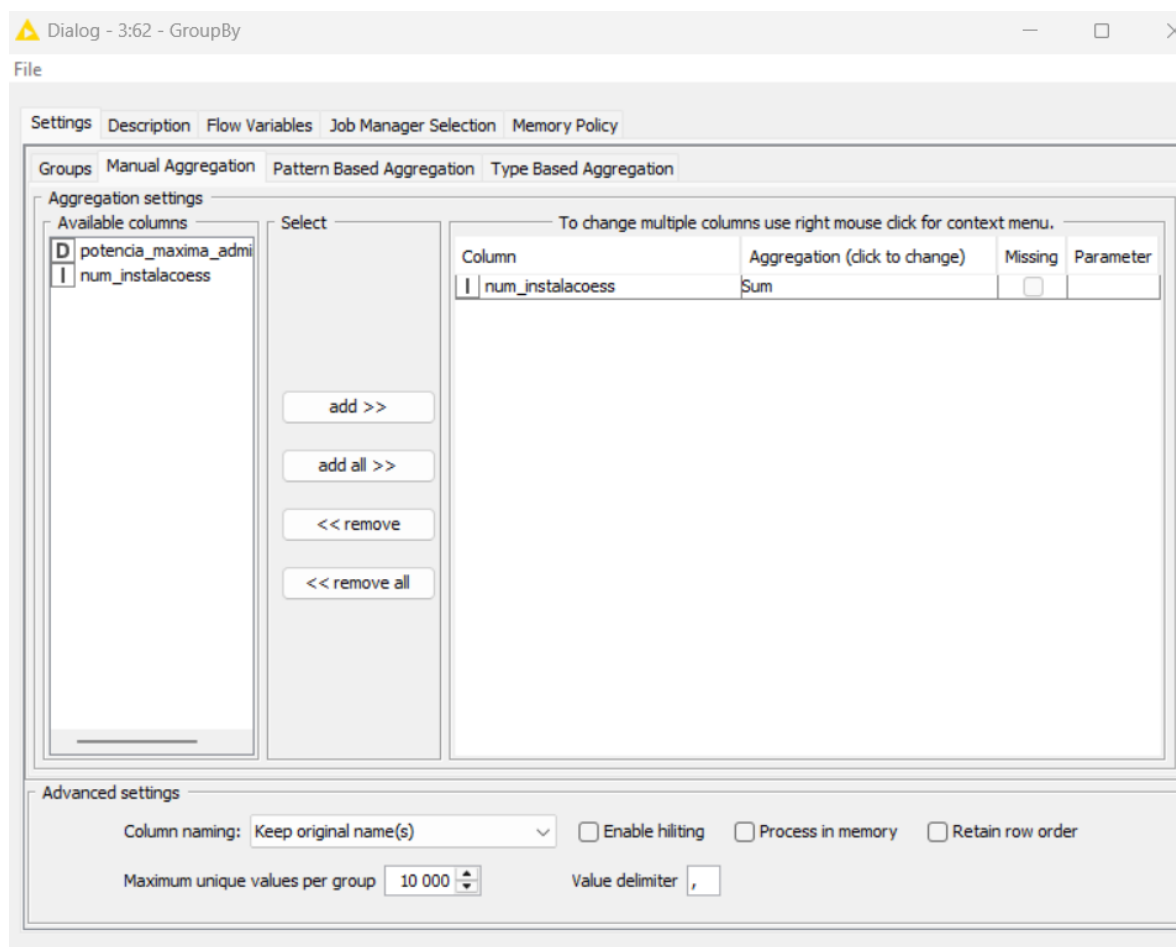


Figura 18 - Configuração Manual GroupBy num_instalacao

Posteriormente, foi utilizado o componente “Joiner” para combinar os dados resultantes dos dois componentes GroupBy numa única tabela.

A junção foi realizada com base no campo “Freguesia”, permitindo integrar as estatísticas da potência máxima dos postos de carregamento com os demais indicadores agregados obtidos anteriormente.

O resultado é uma tabela consolidada, onde cada freguesia contém agora múltiplas métricas relevantes, facilitando comparações diretas e futuras análises integradas.

A Figura 19 ilustra o resultado deste processo de junção dos datasets no KNIME.

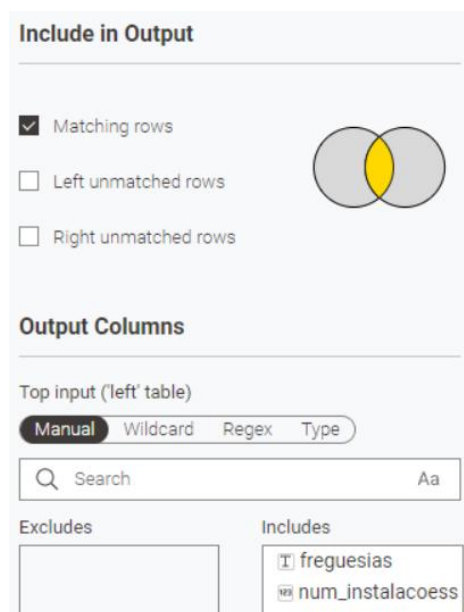


Figura 19 - Configuração do Joiner para os Pontos de Carregamento

Para finalizar o processo de transformação, e já com todos os dados integrados numa única tabela, foi utilizado o componente “Math Formula”. Este componente permitiu calcular a média da potência máxima admissível por posto de carregamento em cada freguesia, criando assim um indicador mais interpretável sobre a eficiência e capacidade média da infraestrutura elétrica local.

A Figura 20 apresenta a fórmula aplicada no KNIME, responsável por efetuar este cálculo para cada freguesia.

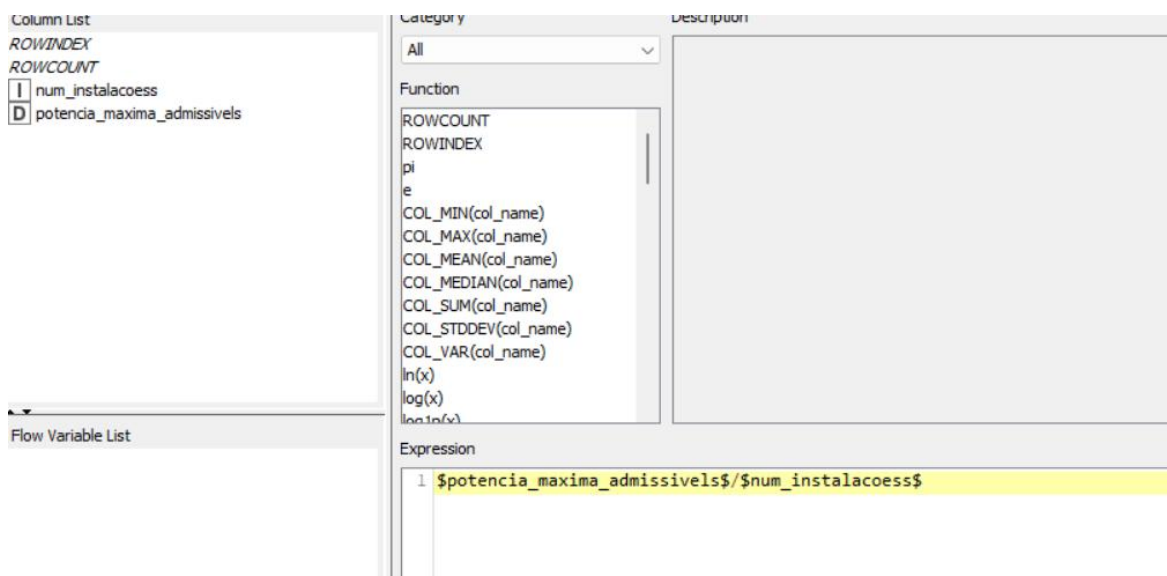


Figura 20 - Configuração Math Formula para Média de Potência por Num Instalação

O resultado é uma tabela na figura 21 com todos os dados úteis transformados prontos para serem usados pelo componente de visualização de dados.

Rows: 13 Columns: 4					Table Statistics	
#	RowID	freguesias	num_instalacoess	potencia_maxima_admissivels	media_por_posto	
1	Row0_J	Alheira e Igreja Nova	3	965.7	321.9	
2	Row1_J	Arcozelo	7	1,040.25	148.607	
3	Row2_J	Barcelinhos	2	250	125	
4	Row3_J	Barcelos, V. Boa, V. Frescainha	14	3,413.6	243.829	
5	Row4_J	Chorente, Góios, Courel, Pedra Furada e Gual	1	3.45	3.45	
6	Row5_J	Fragoso	1	20.7	20.7	
7	Row6_J	Gamil e Midões	1	200	200	
8	Row7_J	Gillmonde	1	120	120	
9	Row8_J	Lijó	1	1,250	1,250	

Figura 21 - Tabela Transformada de Dados Postos Carregamento

Em um dos fluxos do KNIME Workflow, podemos encontrar o componente “Top k Row Filter” este componente é útil para obter as primeiras 10 linhas que cumprem um critério e por uma ordem, como podemos ver na figura 22 o componente e a sua configuração na figura 23.

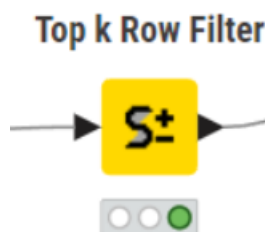


Figura 22 - Componente Top K Row Filter

The image shows the configuration interface for the 'Top K Row Filter' component. The interface is divided into two main sections: 'Sorting' and 'Filter'.

Sorting Section:
 - 'Criterion 1' is displayed with up, down, and delete icons.
 - 'Column' is a dropdown menu currently showing 'cpress'.
 - 'Order' has two buttons: 'Ascending' and 'Descending', with 'Descending' being the selected option.
 - There is a button labeled '+ Add sorting criterion'.
 - A checkbox labeled 'Sort missing values to end' is checked.

Filter Section:
 - 'Filter mode' has two buttons: 'Rows' and 'Unique values', with 'Rows' being the selected option.
 - 'Number of rows / unique values' is a numeric input field currently set to '10'.

Figura 23 - Configuração do Top K Row Filter

Dentro do componente Visualização de Dados podemos ver o diagrama na figura 24, este componente agrega todos os dados transformados no Workflow para visualizar os tais dados em gráficos.

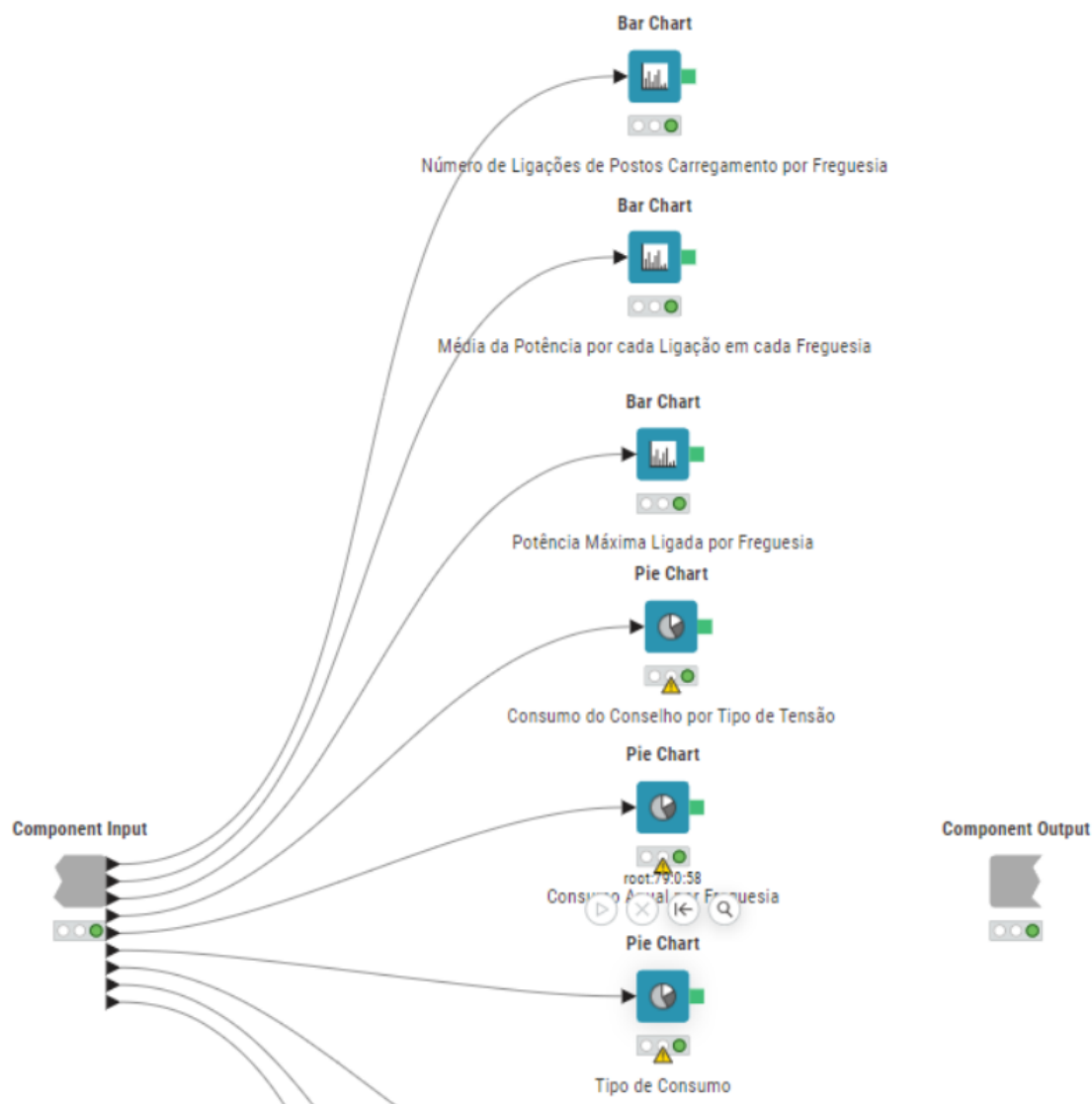


Figura 24 - Diagrama Visualização de Dados

O KNIME contém diversos componentes que permitem gerar gráficos para análise de dados. Neste trabalho, foram selecionados três componentes particularmente úteis para a visualização e interpretação dos resultados. Cada um deles possibilita a criação de gráficos interativos, facilitando a exploração dos dados de forma dinâmica. Na figura 25, podemos observar um exemplo de gráfico interativo gerado por um desses componentes.

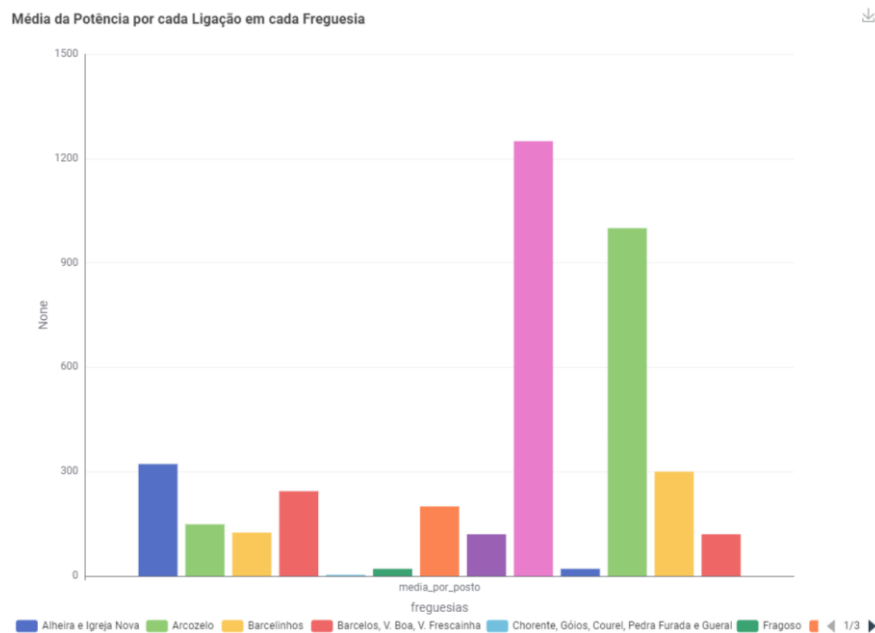


Figura 25 - Gráfico de Barras Média de Potência por Ligação

4. Jobs

No KNIME não existe propriamente a noção de Jobs, como em outras plataformas ETL. Contudo, um workflow do KNIME pode ser entendido como a integração de múltiplas transformações e processos, permitindo criar fluxos de dados automatizados, na figura 26 podemos ver o diagrama de processo de um loop que faz uma chamada GET para cada freguesia de Barcelos, automatizando o processo.

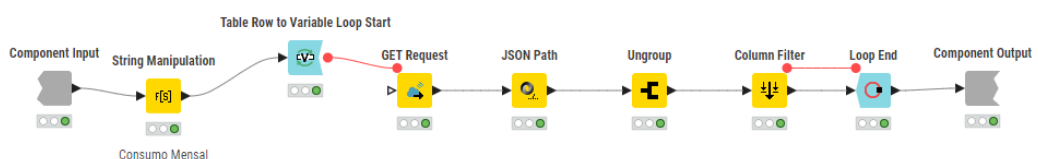


Figura 26 - Diagrama Extração de Dados com Loop

E na componente de visualização dos dados podemos ter a noção de Job em que junta todos os gráficos num dashboard interativo como podemos ver em excerto na figura 27.

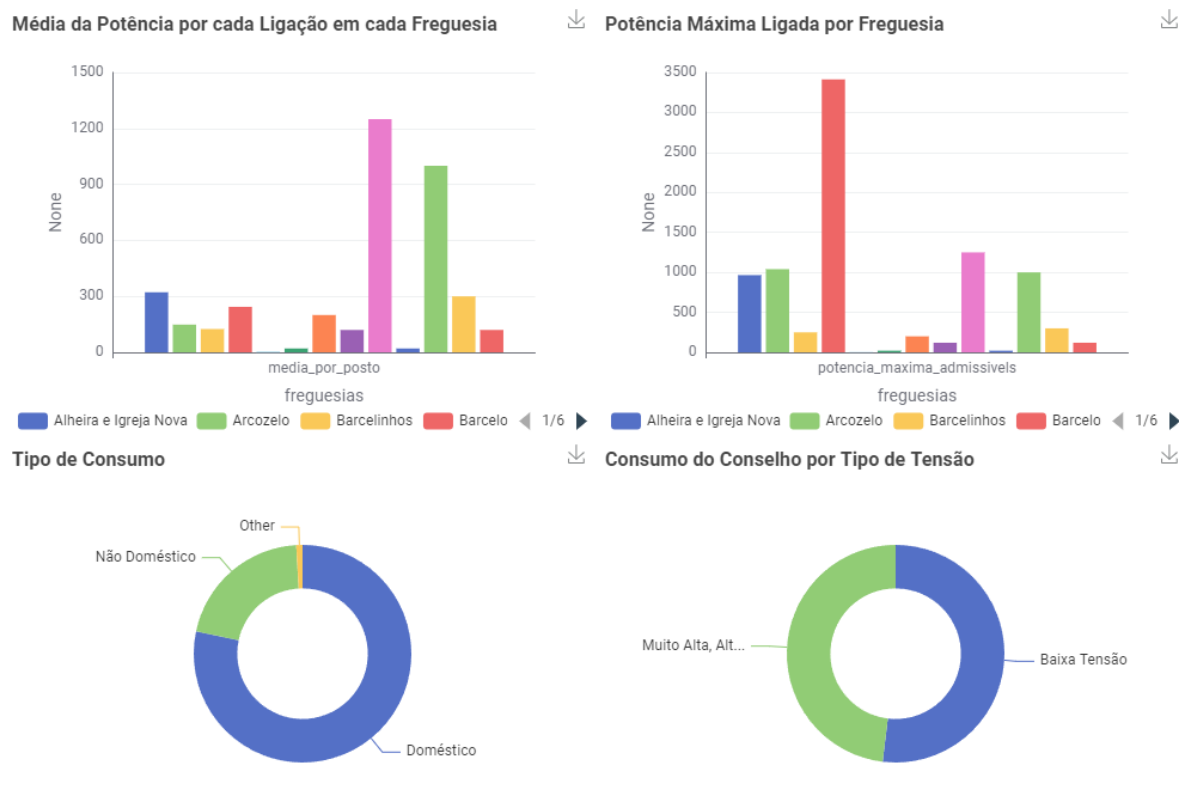


Figura 27 - Excerto Dashboard Interativo

5. Vídeo com demonstração



6. Conclusão

O desenvolvimento deste workflow no KNIME permitiu compreender todo o processo de transformação e análise de dados a partir de ficheiros JSON, até à obtenção de informações e estatísticas úteis para a rede elétrica do concelho de Barcelos.

Através das etapas de tratamento desde a conversão dos dados JSON em tabela, desagrupamento e filtragem, até às operações de agrupamento, junção e cálculo de médias, foi possível estruturar os dados de forma coerente e extraí-los de modo a facilitar a interpretação e visualização.

Os componentes de visualização do KNIME mostraram-se particularmente úteis para representar graficamente os resultados obtidos, permitindo uma análise mais intuitiva e interativa das informações.

O trabalho permitiu consolidar conhecimentos sobre processos ETL, manipulação de dados e visualização.

7. Bibliografia

<https://docs.knime.com/>

<https://e-redes.opendatasoft.com/pages/homepage/>