MYDRAL

Data Summer Course

# HOW TO AUTOMATE YOUR EXCEL TASKS WITH **KNIME ?**

MYDRAL

## AGENDA

01. Analytics platform

02. The context : Building a workflow for preparing data

03. Practical use cases

04. Questions & Answers

MYDRAL

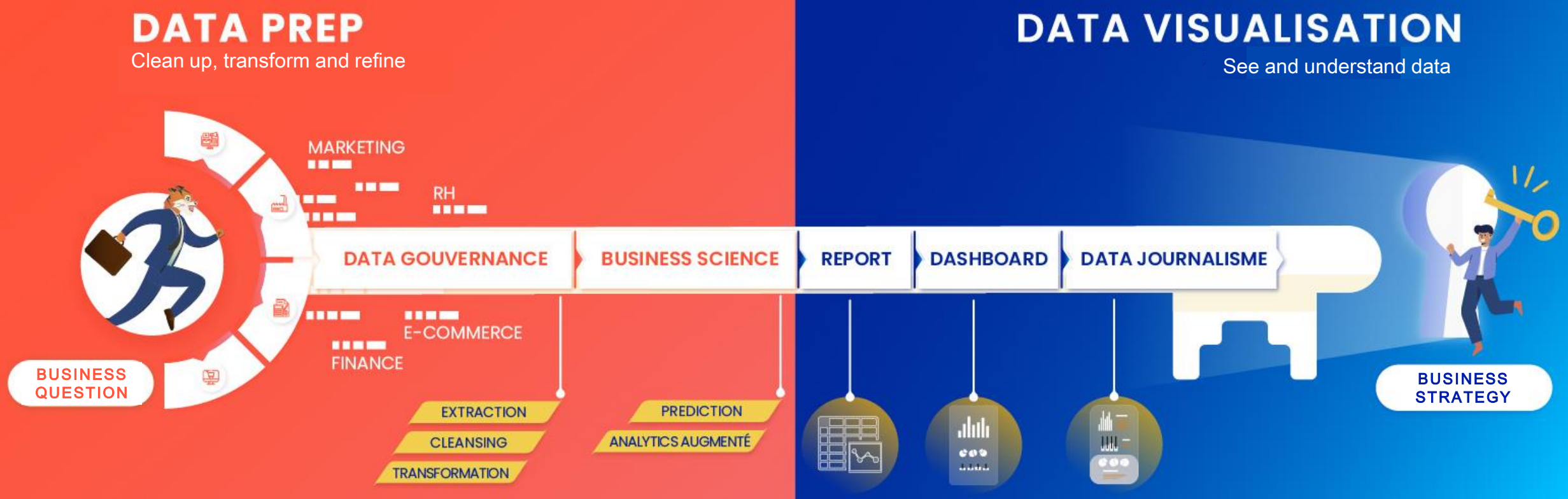
# Analytics Platform

MYDRAL - Data is Power

KNIME – Open to Innovation

MYDRAL

# Data Value Lifecycle



## Who we are

MYDRAL

**PURE PLAYER**

Specialised in generating business value with data

**PROFESSIONAL SERVICE**

Consultancy and support in carrying out data projects

MYDRAL DATA ACADEMY

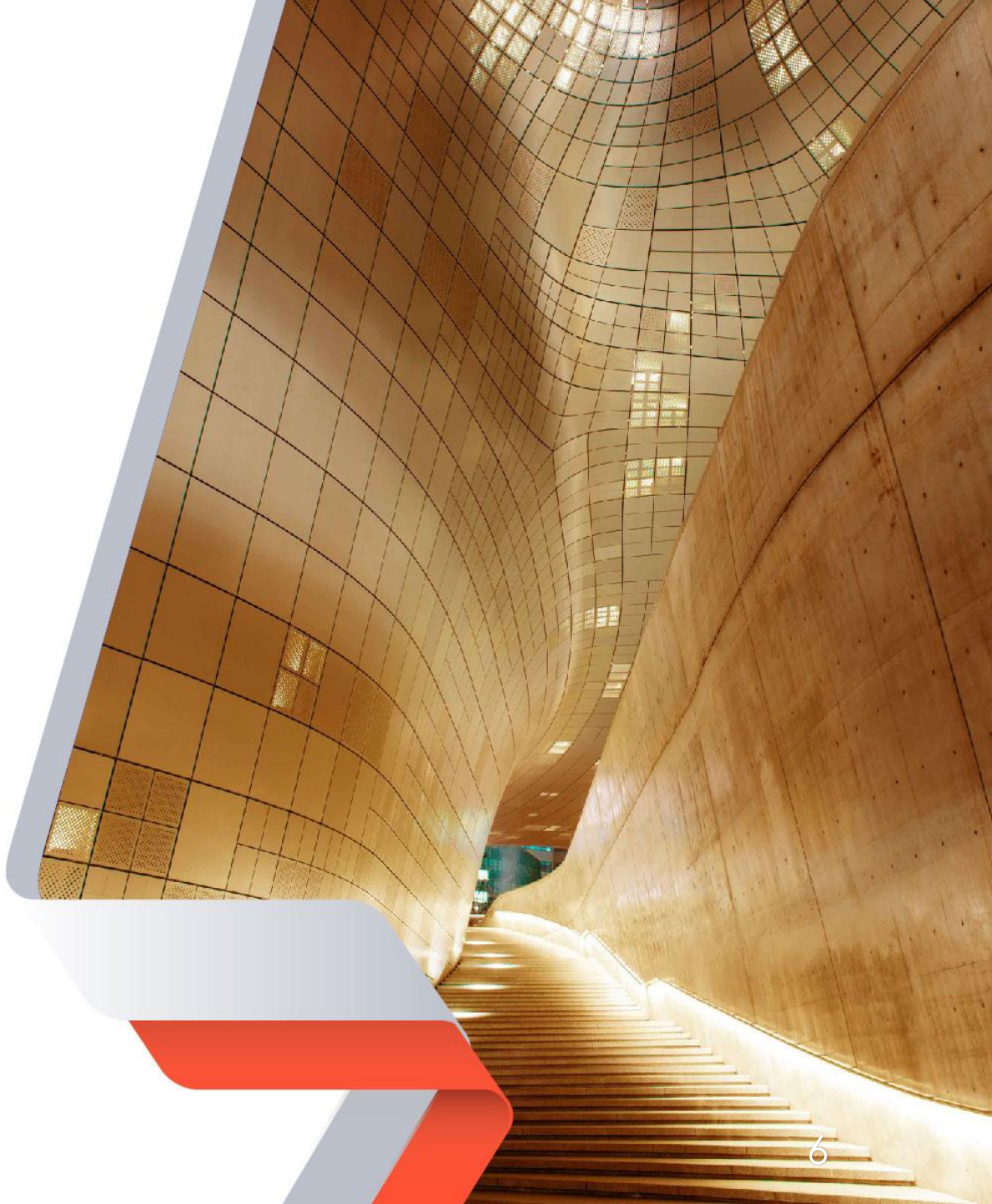
**TRAINING**

4000+ Business Scientists and Analysts trained

MYDRAL

# THE CONTEXT

Building a workflow for preparing data



MYDRAL

## Building a workflow for preparing data

Excel is a very popular business solution for **analyzing data**. However, the use of Excel for data prep needs or more complex tasks can be limited and time consuming

WHY ?

HOW ?

**Transpose** your Excel knowledge with a "workflow" approach to simplify and accelerate the analysis of your data in a more **intuitive way**.

WHAT ?

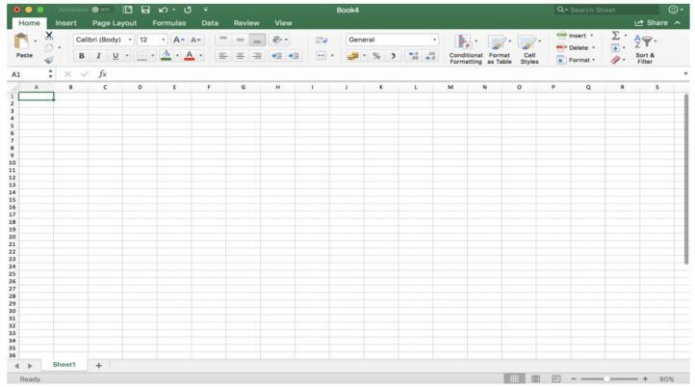
KNIME **simplifies** all your analytical processes and allows you to easily create workflows to **prepare, cleanse and enrich** your data

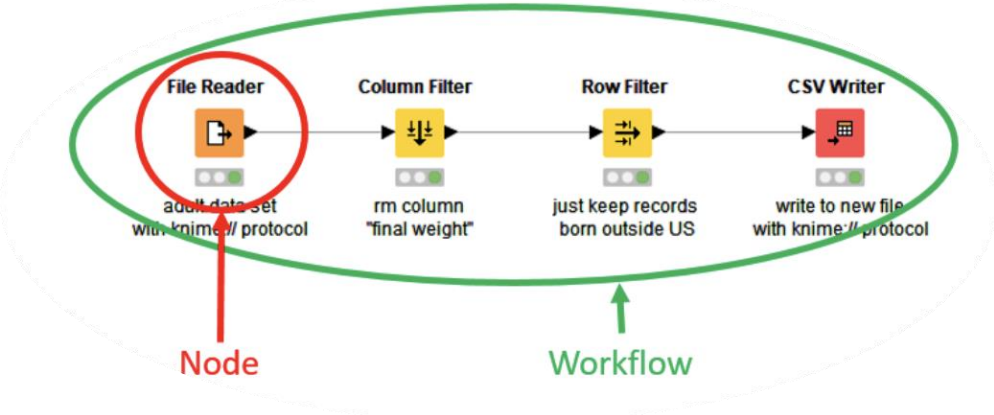
MYDRAL

Transpose your Excel knowledge to KNIME

## What we will cover:

- Fuse
- Connect
- Clean
- Import
- Blend
- Aggregate





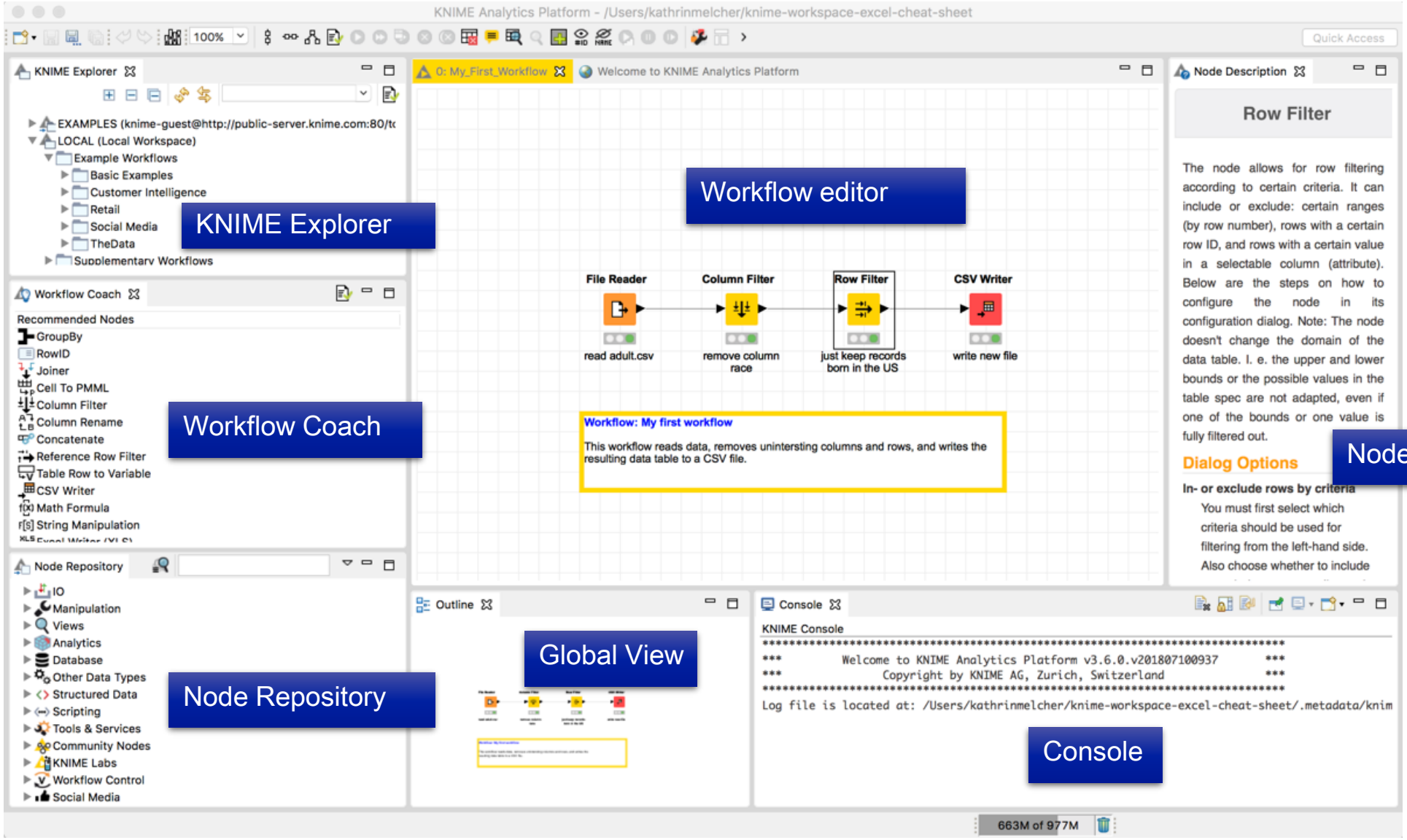
MYDRAL

# PRACTICAL USE CASES

Data preparation with KNIME

MYDRAL

## KNIME – User Interface



MYDRAL

## The data

In this training we will use the information on the total energy consumption at regional and local level in the UK between 2005 and 2017.

Our data source is available here and we will use the geographic data present here

Total sub-national final energy consumption, 2005

| | | Coal [2] | | | | Manufactured fuels [3] | | | Petroleum products [2] | | | | | | | Gas [6] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LA Code | Government Office Regions and LAU1 Areas | Industrial & Commercial | Domestic | Rail | Total | Industrial | Domestic | Total | Industrial & Commercial | Domestic | Road transport | Rail | Public Sector | Agriculture | Total | Industrial & Commercial | Domestic | T |
| W06000019 | Blaenau Gwent | 0,5 | 1,1 | - | **1,5** | 0,1 | 0,7 | **0,7** | 7,0 | 0,4 | 25,4 | 0,1 | 0,1 | 0,1 | **33,1** | 20,1 | 55,9 | |
| W06000013 | Bridgend | 1,7 | 1,5 | - | **3,2** | 14,0 | 1,0 | **15,1** | 13,3 | 1,9 | 89,5 | 2,0 | 0,2 | 1,8 | **108,8** | 70,9 | 94,0 | |
| W06000018 | Caerphilly | 1,6 | 3,0 | - | **4,6** | 0,6 | 1,8 | **2,5** | 16,6 | 1,7 | 72,2 | 1,2 | 0,2 | 1,6 | **93,5** | 45,5 | 122,7 | |
| W06000015 | Cardiff | 2,2 | 1,3 | - | **3,5** | 14,4 | 1,4 | **15,8** | 16,2 | 1,1 | 210,8 | 2,7 | 0,7 | 0,8 | **232,3** | 222,6 | 199,4 | |
| W06000010 | Carmarthenshire | 1,8 | 7,2 | 0,0 | **9,1** | 6,5 | 4,1 | **10,5** | 26,1 | 46,4 | 119,9 | 1,4 | 0,7 | 30,7 | **225,2** | 62,3 | 76,7 | |
| W06000008 | Ceredigion | 1,0 | 4,0 | 0,1 | **5,1** | 0,0 | 2,3 | **2,3** | 33,7 | 29,2 | 42,5 | 0,8 | 1,0 | 19,7 | **126,9** | 8,3 | 12,0 | |
| W06000003 | Conwy | 1,2 | 2,1 | - | **3,3** | 0,2 | 1,4 | **1,6** | 10,8 | 9,0 | 77,2 | 0,9 | 0,4 | 5,0 | **103,4** | 21,1 | 67,3 | |
| W06000004 | Denbighshire | 2,0 | 2,3 | 0,0 | **4,4** | 0,0 | 1,4 | **1,5** | 13,9 | 12,4 | 57,4 | 0,8 | 0,3 | 5,8 | **90,5** | 21,0 | 49,6 | |
| W06000005 | Flintshire | 24,0 | 2,5 | - | **26,5** | 13,1 | 1,6 | **14,7** | 45,9 | 17,1 | 116,8 | 1,0 | 0,3 | 4,1 | **185,3** | 86,2 | 79,0 | |
| W06000002 | Gwynedd | 2,4 | 5,2 | 0,3 | **7,9** | 0,0 | 3,2 | **3,2** | 37,4 | 27,5 | 79,8 | 1,3 | 0,8 | 12,4 | **159,2** | 22,9 | 39,8 | |
| W06000001 | Isle of Anglesey | 0,8 | 2,1 | - | **2,9** | 0,5 | 1,4 | **1,9** | 9,4 | 19,6 | 39,1 | 0,6 | 0,3 | 9,8 | **78,8** | 27,3 | 21,9 | |
| W06000024 | Merthyr Tydfil | 0,1 | 0,9 | 0,0 | **1,0** | 0,1 | 0,6 | **0,6** | 3,6 | 0,9 | 27,0 | 0,9 | 0,0 | 0,3 | **32,6** | 18,7 | 43,9 | |
| W06000021 | Monmouthshire | 0,8 | 2,0 | - | **2,8** | 0,0 | 1,2 | **1,3** | 7,5 | 13,2 | 100,7 | 1,8 | 0,2 | 8,8 | **132,2** | 39,1 | 44,9 | |
| W06000012 | Neath Port Talbot | 44,2 | 3,0 | - | **47,2** | 538,8 | 1,8 | **540,6** | 57,7 | 4,7 | 91,0 | 2,3 | 0,2 | 2,0 | **157,9** | 37,2 | 92,8 | |
| W06000022 | Newport | 0,3 | 1,0 | - | **1,2** | 9,1 | 0,8 | **9,9** | 21,4 | 2,8 | 137,4 | 2,1 | 0,2 | 1,8 | **165,7** | 136,5 | 89,5 | |
| W06000009 | Pembrokeshire | 2,5 | 3,1 | - | **5,7** | 273,9 | 2,0 | **275,9** | 1 060,6 | 31,7 | 63,7 | 0,8 | 0,5 | 20,0 | **1 177,3** | 11,8 | 47,4 | |
| W06000023 | Powys | 2,3 | 6,4 | 0,1 | **8,8** | 0,8 | 3,6 | **4,5** | 30,2 | 44,0 | 102,2 | 0,6 | 0,9 | 31,6 | **209,5** | 19,0 | 40,4 | |
| W06000016 | Rhondda Cynon Taf | 1,8 | 3,9 | - | **5,7** | 0,5 | 2,5 | **3,0** | 20,1 | 1,7 | 137,4 | 2,6 | 0,2 | 2,0 | **164,0** | 63,3 | 168,1 | |
| W06000011 | Swansea | 0,4 | 2,6 | - | **3,0** | 0,7 | 1,9 | **2,5** | 26,4 | 6,6 | 116,0 | 1,0 | 0,3 | 3,0 | **153,2** | 90,4 | 154,1 | |
| W06000020 | Torfaen | 0,3 | 0,9 | 0,0 | **1,2** | 0,1 | 0,6 | **0,8** | 10,8 | 0,7 | 41,1 | 0,5 | 0,1 | 0,5 | **53,6** | 42,6 | 62,0 | |
| W06000014 | Vale of Glamorgan | 22,1 | 1,1 | 0,0 | **23,2** | 18,1 | 0,8 | **18,9** | 14,3 | 5,5 | 68,3 | 2,3 | 0,2 | 3,8 | **94,5** | 42,8 | 76,7 | |
| W06000006 | Wrexham | 3,4 | 2,0 | - | **5,4** | 0,1 | 1,4 | **1,4** | 18,9 | 9,8 | 66,4 | 0,7 | 0,2 | 4,5 | **100,5** | 130,7 | 71,8 | |
| **W92000004** | **WALES** | 117,4 | 59,2 | 0,5 | **177,1** | 891,6 | 37,5 | **929,2** | 1 501,7 | 287,9 | 1 881,9 | 28,4 | 8,3 | 169,9 | **3 878,0** | 1 240,1 | 1 710,1 | |
| S12000033 | Aberdeen City | 1,2 | 0,4 | - | **1,6** | 1,0 | 0,4 | **1,4** | 59,7 | 2,3 | 94,5 | 0,5 | 2,3 | 1,0 | **160,3** | 125,8 | 144,7 | |
| S12000034 | Aberdeenshire | 3,8 | 4,8 | - | **8,6** | 1,3 | 2,9 | **4,2** | 59,7 | 65,4 | 176,9 | 1,6 | 1,1 | 31,2 | **336,0** | 59,6 | 98,0 | |
| S12000041 | Angus | 1,7 | 1,3 | 0,0 | **3,0** | - | 0,8 | **0,8** | 20,7 | 14,0 | 68,4 | 0,9 | 0,1 | 7,4 | **111,6** | 30,3 | 60,1 | |
| S12000035 | Argyll and Bute | 1,2 | 3,2 | - | **4,4** | 0,0 | 1,7 | **1,8** | 32,8 | 15,0 | 57,4 | 0,5 | 6,6 | 10,3 | **122,5** | 14,5 | 35,1 | |
| S12000005 | Clackmannanshire | 0,1 | 0,3 | - | **0,4** | - | 0,2 | **0,2** | 4,5 | 1,0 | 19,1 | 4,2 | 0,0 | 0,9 | **29,7** | 41,2 | 34,1 | |
| S12000006 | Dumfries and Galloway | 2,2 | 3,5 | - | **5,8** | 0,0 | 2,0 | **2,0** | 34,7 | 30,1 | 164,5 | 4,1 | 1,0 | 43,5 | **277,8** | 50,6 | 65,8 | |
| S12000042 | Dundee City | 0,0 | 0,3 | - | **0,3** | 0,1 | 0,3 | **0,5** | 22,4 | 0,5 | 62,7 | 0,3 | 0,1 | 0,1 | **86,1** | 71,0 | 79,6 | |
| S12000008 | East Ayrshire | 0,5 | 0,7 | - | **1,3** | 0,0 | 0,5 | **0,5** | 13,4 | 4,3 | 78,3 | 2,1 | 0,5 | 8,4 | **107,0** | 41,7 | 85,1 | |
| S12000045 | East Dunbartonshire | 0,0 | 0,2 | - | **0,2** | 0,0 | 0,2 | **0,3** | 4,8 | 0,9 | 41,1 | 4,5 | 0,1 | 0,8 | **52,2** | 17,2 | 85,7 | |
| S12000010 | East Lothian | 25,9 | 1,3 | - | **27,2** | 21,2 | 0,7 | **21,9** | 9,3 | 5,8 | 62,0 | 0,2 | 0,4 | 3,3 | **81,0** | 18,8 | 56,9 | |
| S12000011 | East Renfrewshire | 0,0 | 0,2 | - | **0,2** | 0,0 | 0,2 | **0,2** | 2,2 | 1,0 | 58,5 | 1,3 | 0,1 | 1,3 | **64,5** | 12,9 | 70,4 | |
| S12000036 | City of Edinburgh | 0,1 | 0,8 | - | **0,9** | 0,0 | 0,8 | **0,8** | 23,0 | 2,0 | 243,5 | 2,6 | 1,5 | 0,9 | **273,4** | 201,6 | 295,4 | |
| S12000013 | Na h-Eileanan Siar | 1,2 | 2,3 | - | **3,5** | - | 1,2 | **1,2** | 15,4 | 15,0 | 13,2 | 0,5 | 0,6 | 3,1 | **47,9** | - | - | |

MYDRAL

## Steps

Excel Files

Building Workflow

Data Visualisation

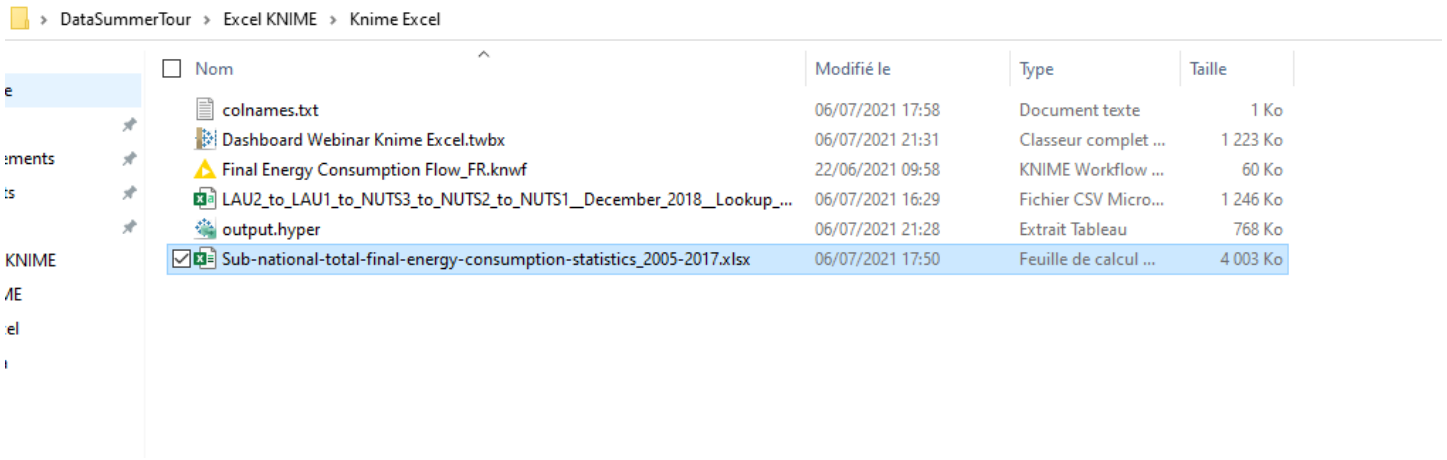
MYDRAL

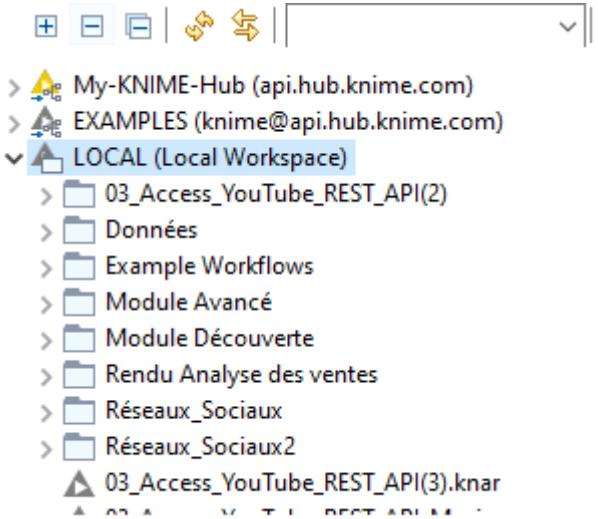
# Introduction – Opening the workflow

Preparing data with KNIME

# Opening the workflow

- The files we are going to use is called « Sub-national-total-final-energy-consumption-statistics_2005-2017 ». To use it we will first create a Knime workflow




- To create a Knime Workflow you have to right click on the "Local" folder then on "New Knime Workflow"

MYDRAL

# Module 01 – Import data

Preparing data with KNIME

# Importation des données

- Preparing and cleaning data is necessary before any analysis..

- To make a sales analysis, we need to extract the list of orders. The IT department gives us access to the following Excel files…

  Sub-national-total-final-energy-consumption-statistics_2005-2017.xlsx

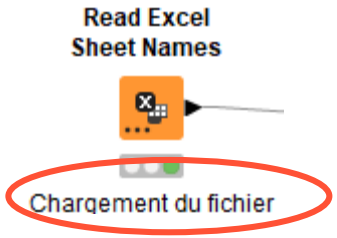
  … which includes information for the last 12 months.

- We are going to create a processing workflow to prepare the data.

# Import data

- We will start by using the “Read Excel Sheet Name” node to load the names of the different sheets. As a reminder, the file to import into this Node is the "Sub-national-total-final-energy-consumption-statistics_2005-2017.xlsx".

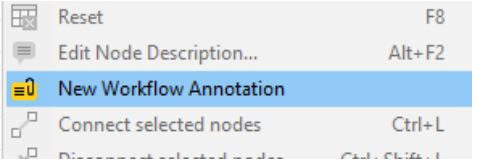
In order to comment on the Nodes and know exactly what they refer to, you must double click on the pre-written comment “Node X”



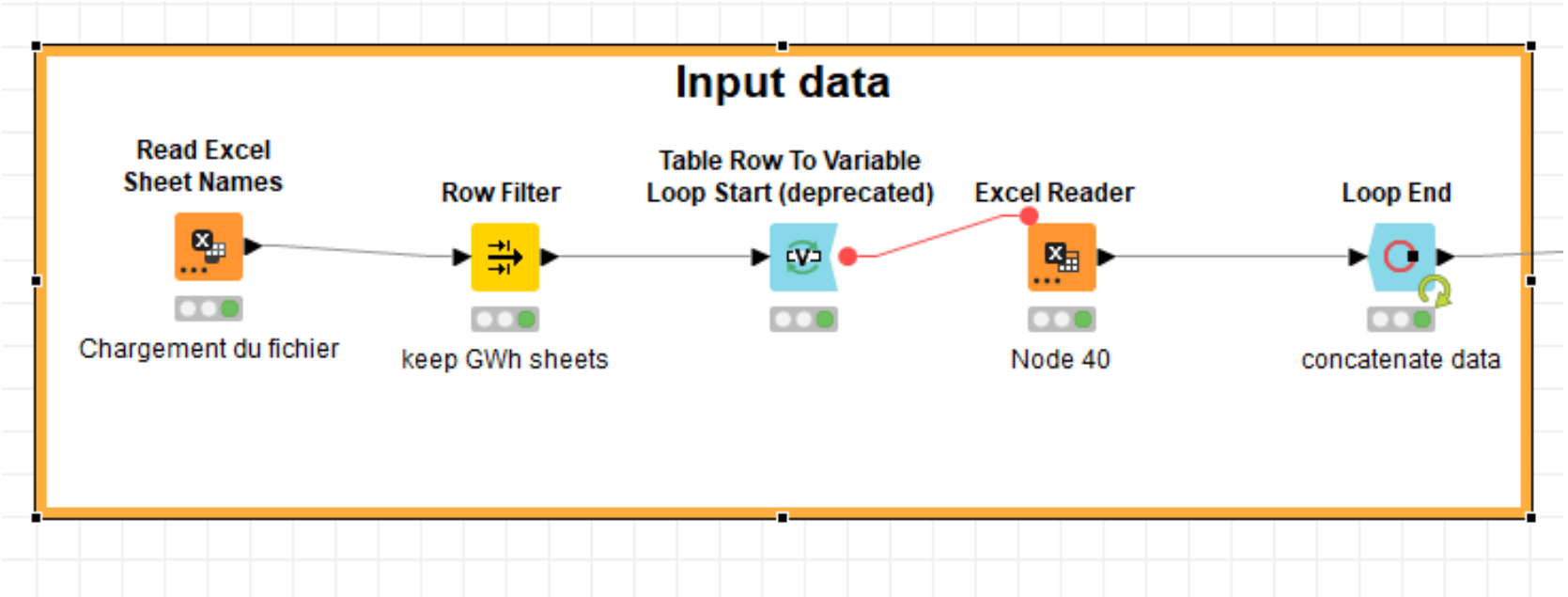

Expected result :

# Import data

- To get our bearings in our workflow, we will start by creating an annotation that will allow us to put a title to our step. We do a right click then:




- You can then click on the pen at the top left of the annotation and double click to change its color, size, give it a title (here « *Input Data* ») ...



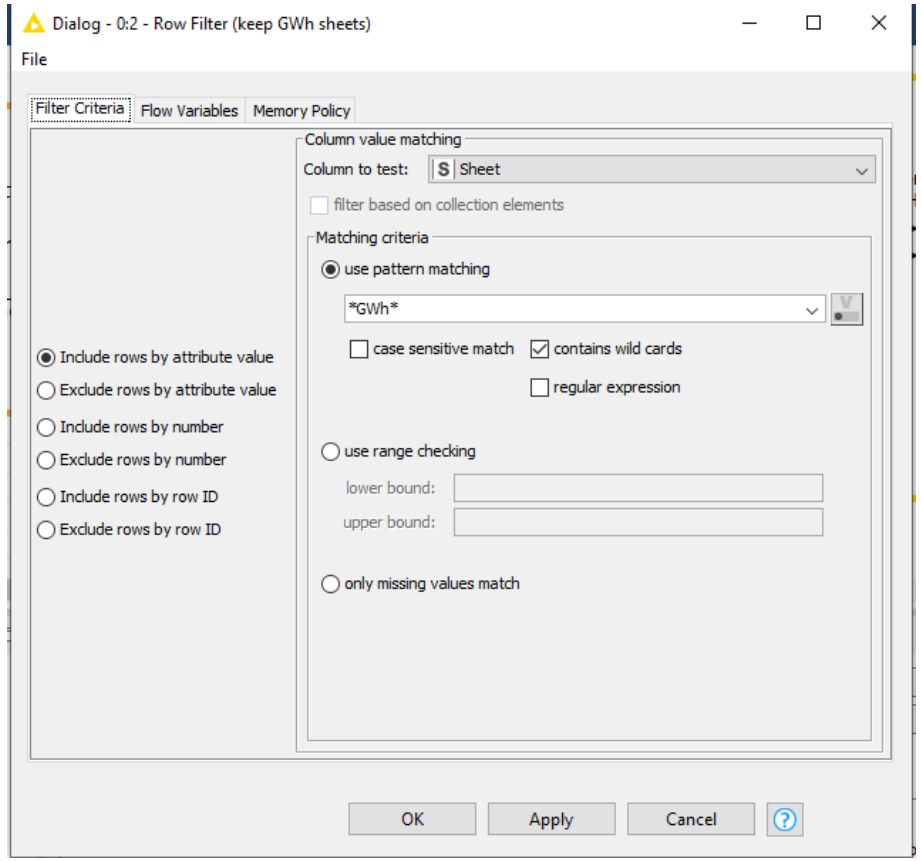
MYDRAL

# Module 02 – Action loop

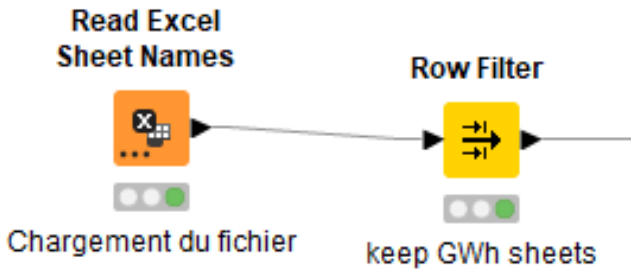
Preparing data with KNIME

# Action loop

- We will then use the "Row Filter" node which allows us to select the rows that we think are useful for the analysis. (here: *GWh*)
- We will use the Pattern Matching option to keep only the lines that correspond to our search




Expected Results :

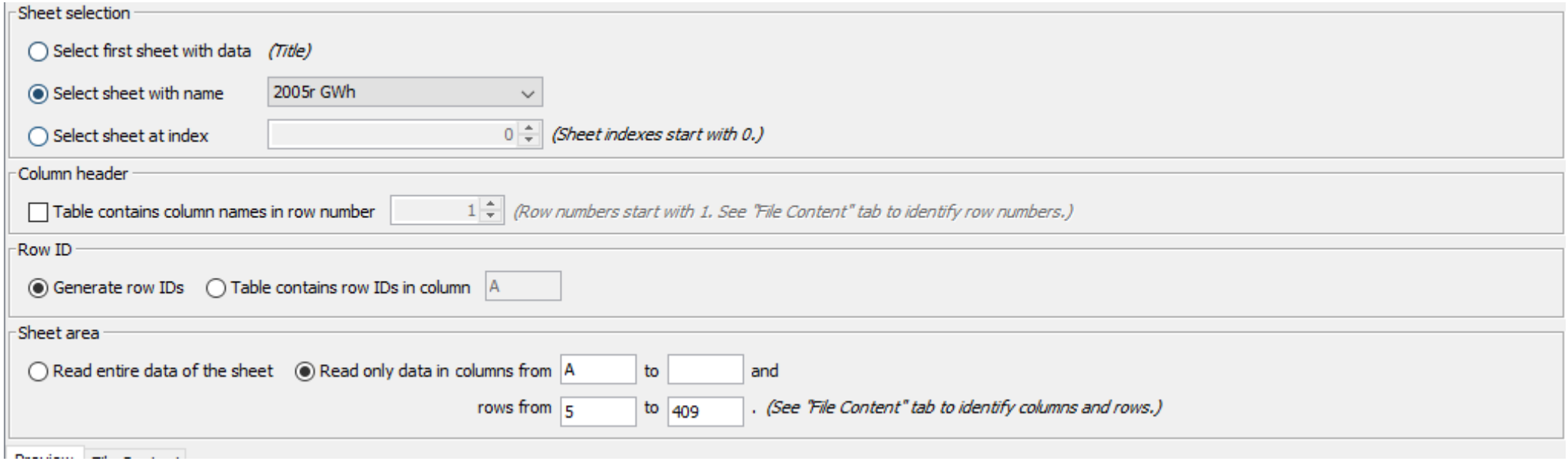


# Action loop

- We are going to create a loop in order to retrieve the information from the different sheets of our excel file in order to concatenate them together.

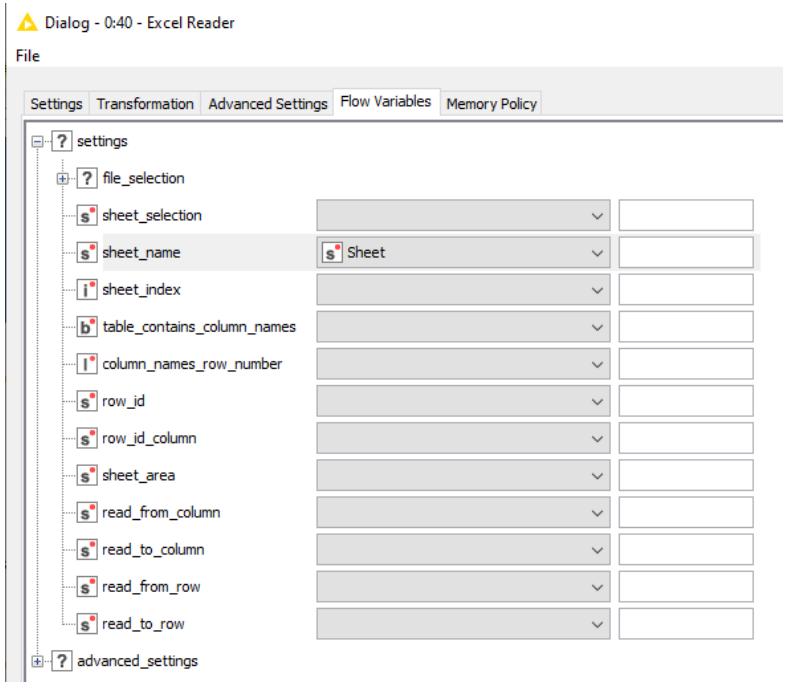
- We select the "Table Row To Variable Loop Start" node and drag it into the workspace. There will be no settings to change.

- We then select the "Excel Reader" node and drag it into the workspace. We double click to open the Node and we add the file "Sub-national-total-final-energy-consumption-statistics_2005-2017.xlsx".

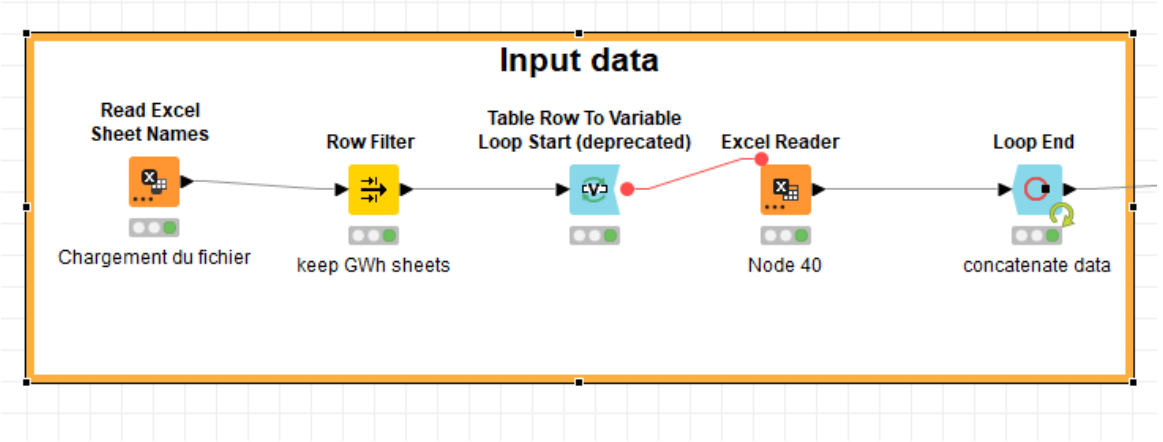


# Action loop

- Then click on the "Flow Variable" tab then on "Settings" and in "Sheet names" select the "Sheet" parameter.




Expected Results :




- Add the Node « Loop End ».

MYDRAL

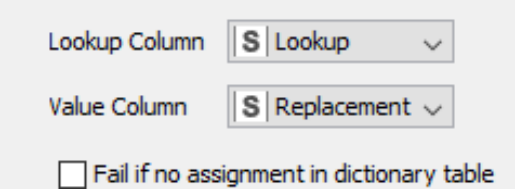
# Module 03 – Cleaning data

Preparing data with KNIME

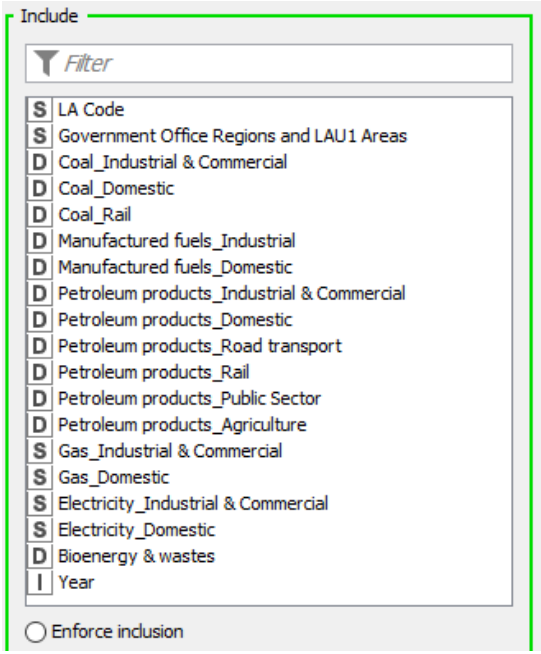
# Cleaning data

- We are going to use the “File Reader” Node to read the “colnames.txt” file including the new names of our columns

- We will also use the "Insert Column Header" node which allows you to replace one line by another. Here we will replace the default titles with custom titles.

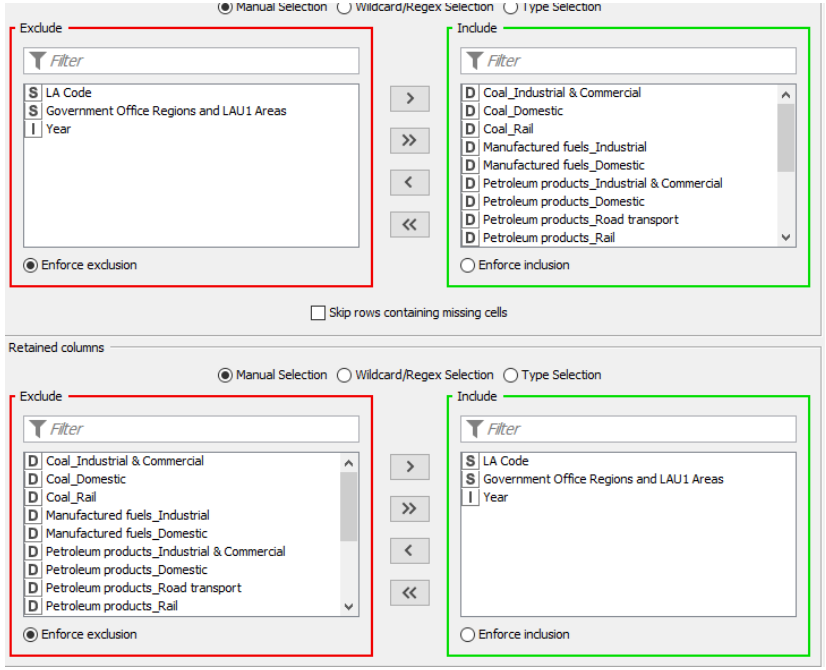



- We add the “Column filter” node to keep only the information relevant to our analysis.

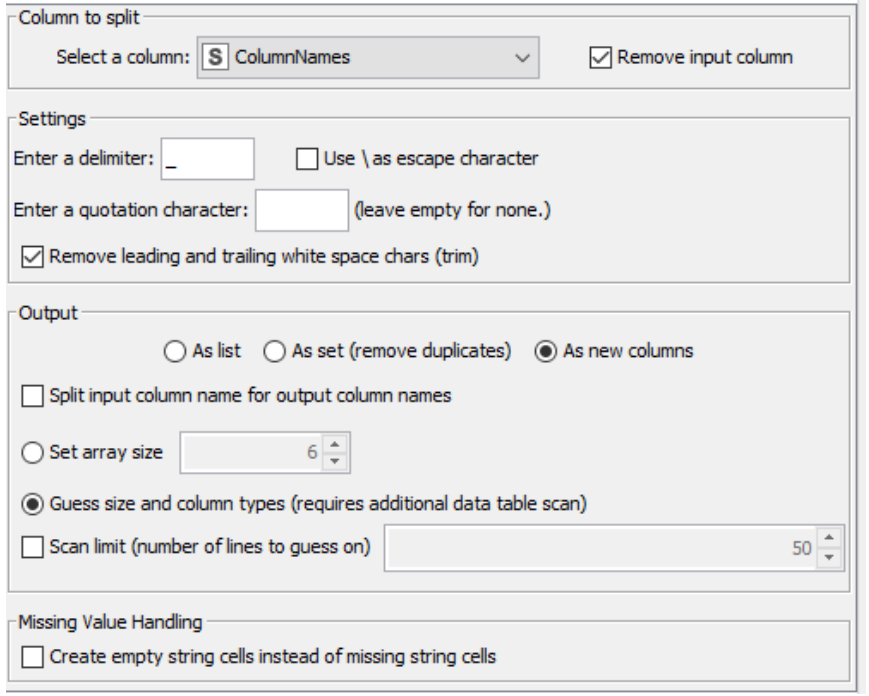


# Cleaning data

- We will then use the “Unpivoting” Node to change the format of our table to database format. The Node rotates the elements of our table according to the parameters that we enter.




- We place the Node "Cell Splitter" in order to categorize our data



# Cleaning data

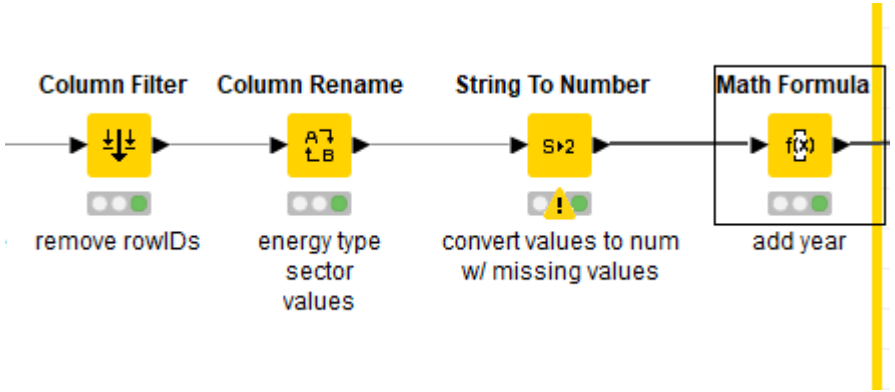
- We place the "Column Filter" node in order to remove the columns linked to the IDs and we rename the column titles with the "Column rename" node. We rename the columns "Column value, ColumnNames_Arr [0], ColumnNames_Arr [1]" by "GWh Value, Energy Type, Sector"

- The Node « **String to Number** » allows you to set the years of type String to Number.

- Finally we place the Node « **Math Formula** » to enter the following formula:

**$Year$+2005**

Expected result:



MYDRAL

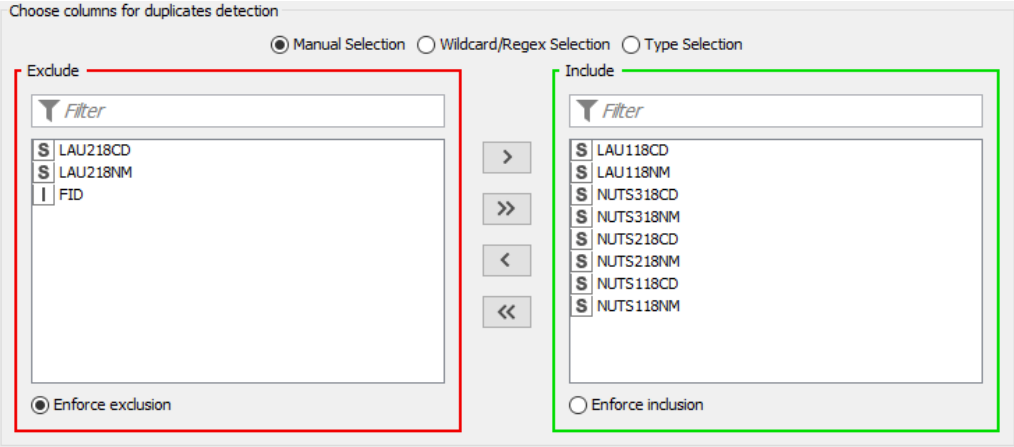
# Module 04 – Enrich Data

Preparing data with KNIME

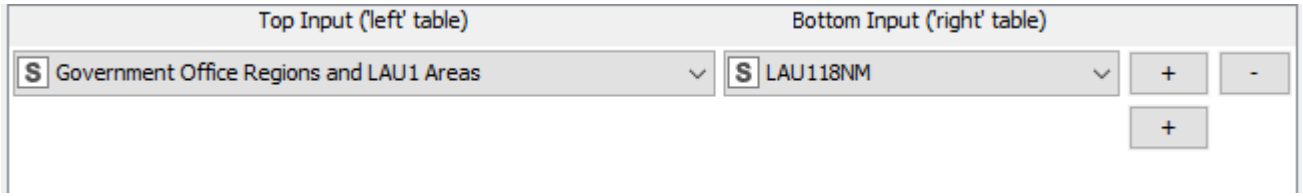
# Enrich Data

- Here, we add a table to enrich our data. This will make the analysis more interesting. We use the "File Reader" node to import the table «LAU2_to_LAU1_to_NUTS3_to_NUTS2_to_NUTS1__December_2018__Lookup_in_United_Kingdom.csv».

- We then use the Node "Duplicate Row Filter" which makes it possible to clean the table of its duplicates.



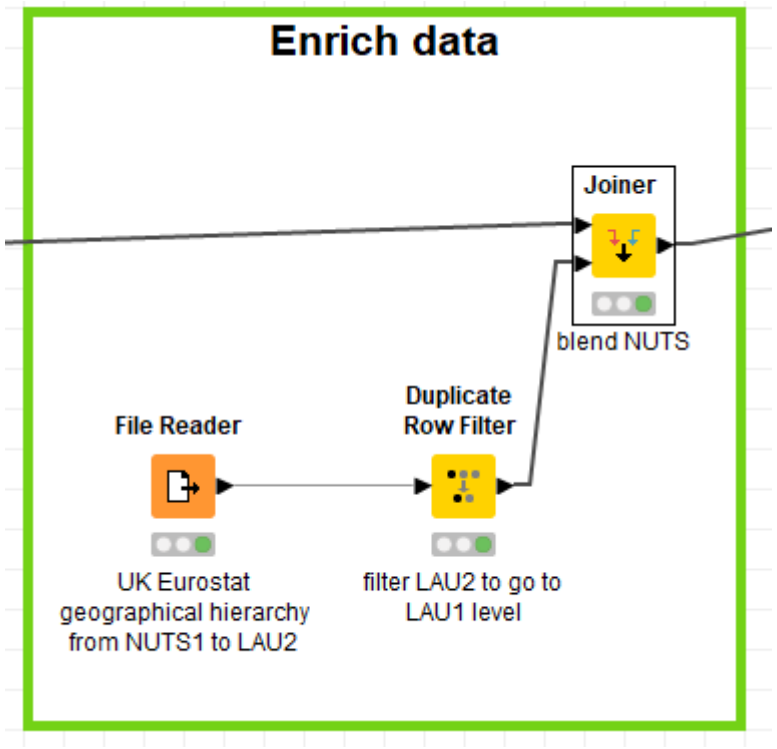

- We end this part with a Node "Joiner" to make the connection between our data sources.



# Enrich Data



Expected result:



MYDRAL

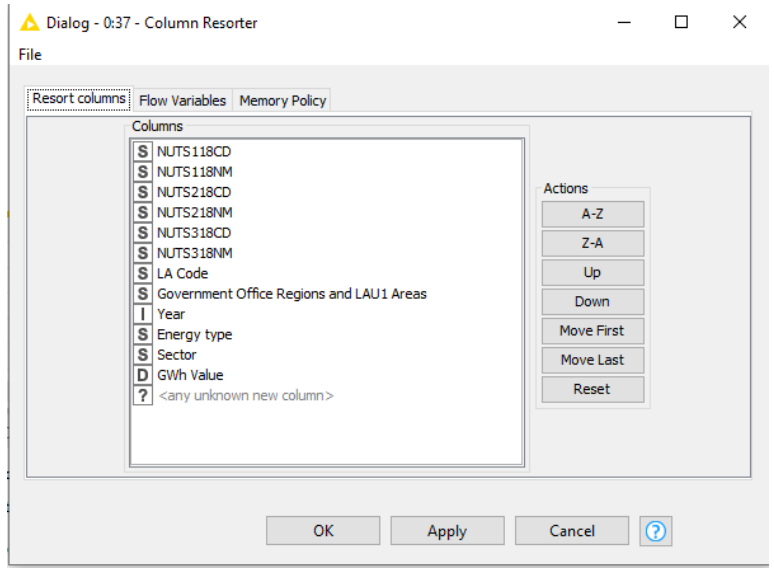
# Module 05 – Exporting Data

Preparing data with KNIME

# Exporting data

- We can start by creating an “Output” annotation in order to locate ourselves correctly in the workflow.

- We then select the "Column Resorter" node to change the order of the columns in our table




- Finally, we choose the "Tableau Writer" node to facilitate the use of our database by Tableau. We could also have used the Node "Excel Writer"

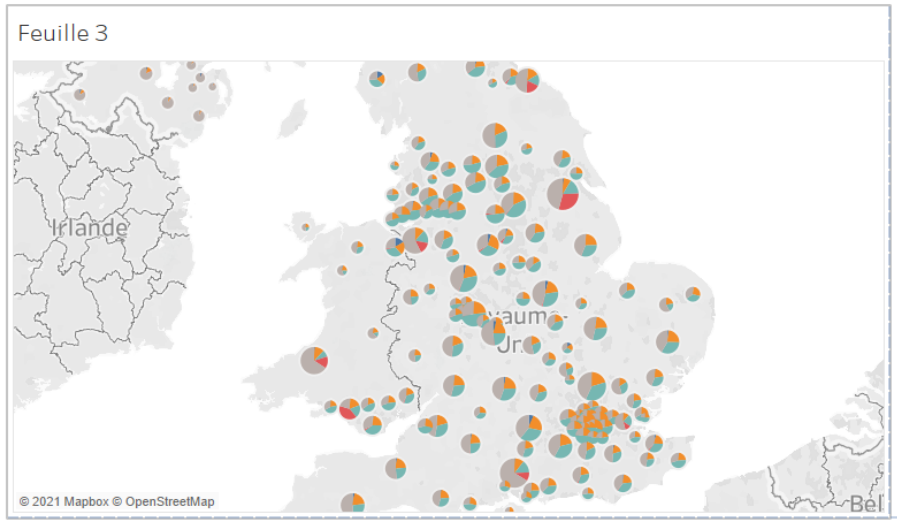
MYDRAL

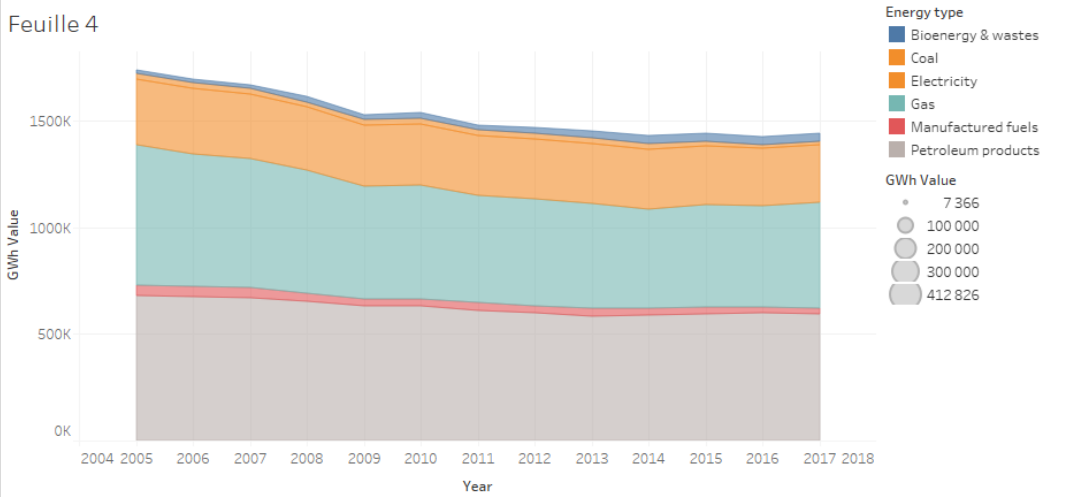
# Module 06 – The Dashboard

Preparing data with KNIME

# The Dashboard

- We can now connect our data source to a Tableau Dashboard and start mining the data:




Feuille 5

| | Energy type / Sector | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bioenergy & wastes | | Coal | | | | Electricity | | | Gas | | | Manufactured fuels | |
| Government Office.. | Null | Total | Domestic | Industrial & Co.. | Rail | Total | Domestic | Industrial & Co.. | Total | Domestic | Industrial & Co.. | Total | Domestic | Industrial |
| Aberdeen City | 442 | 442 | 67 | 174 | 0 | 241 | 5 997 | 10 620 | 16 617 | 18 953 | 13 454 | 32 407 | 50 | 213 |
| Aberdeenshire | 1 127 | 1 127 | 725 | 913 | 0 | 1 638 | 8 038 | 8 969 | 17 006 | 14 042 | 4 980 | 19 021 | 377 | 233 |
| Adur | 373 | 373 | 54 | 1 | 0 | 54 | 1 507 | 1 310 | 2 817 | 4 263 | 1 039 | 5 302 | 44 | 16 |
| Allerdale | 13 528 | 13 528 | 415 | 156 | 0 | 572 | 2 449 | 5 531 | 7 979 | 7 415 | 10 635 | 18 050 | 206 | 316 |
| Amber Valley | 407 | 407 | 194 | 562 | 29 | 785 | 2 902 | 5 035 | 7 937 | 10 540 | 5 247 | 15 788 | 104 | 64 |
| Angus | 413 | 413 | 191 | 376 | 0 | 567 | 3 581 | 3 846 | 7 426 | 7 886 | 4 485 | 12 371 | 105 | 0 |
| Antrim and Newto.. | 1 203 | 1 203 | 395 | 1 770 | 0 | 2 165 | | | | | | | 232 | 80 |
| Ards and North Do.. | 1 363 | 1 363 | 514 | 328 | 0 | 842 | | | | | | | 299 | 9 |
| Armagh City, Banb.. | 1 849 | 1 849 | 893 | 2 236 | 0 | 3 129 | | | | | | | 484 | 19 |
| Arun | 919 | 919 | 142 | 65 | 0 | 207 | 3 824 | 2 996 | 6 820 | 11 648 | 4 051 | 15 699 | 112 | 4 |
| Ashfield | 238 | 238 | 208 | 21 | 0 | 229 | 2 568 | 4 906 | 7 474 | 10 092 | 3 894 | 13 986 | 105 | 2 |
| Ashford | 1 137 | 1 137 | 185 | 534 | 23 | 741 | 2 952 | 3 716 | 6 667 | 6 984 | 3 038 | 10 022 | 121 | 10 |
| Aylesbury Vale | 1 022 | 1 022 | 216 | 410 | 0 | 626 | 4 499 | 4 208 | 8 707 | 11 366 | 4 477 | 15 843 | 149 | 11 |
| Babergh | 360 | 360 | 145 | 407 | 0 | 552 | 2 507 | 2 930 | 5 437 | 4 538 | 2 051 | 6 589 | 82 | 4 |
| Barking and Dagen.. | 456 | 456 | 21 | 0 | 0 | 22 | 3 656 | 5 676 | 9 332 | 10 669 | 3 072 | 13 741 | 50 | 19 |
| Barnet | 423 | 423 | 20 | 2 | 0 | 23 | 8 413 | 6 983 | 15 395 | 30 781 | 7 912 | 38 693 | 48 | 49 |

# Get in touch !

- ✓ You have completed your course, well done !
- ✓ **What is the next step ?** You will receive a link by email to connect with Mydral experts on Thursday August 26 for Q&A and more ! We are here to help !



## Contact Julien NORDMAN

yehouman@mydral.com


## Or via Mydral website :

https://www.mydral.com/en/homepage/

Follow **@MydralUK**

MYDRAL

# Thank you !

# DATA IS POWER