

from Excel® to KNINE®

Author: Kathrin Melcher

Copyright © 2022 by KNIME Press

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording or likewise.

This book has been updated for KNIME 4.6.

For information regarding permissions and sales, write to:

KNIME Press Talacker 50 8001 Zurich Switzerland

knimepress@knime.com

Table of Contents

General Usage	5
Spreadsheets Workflows and Nodes	6
Folders Workspace	7
The KNIME Workbench	
Building a KNIME Workflow	9
Display Data Table	
Input/Output	
Opening an Excel File Excel Reader Node	
Opening a CSV or txt File CSV Reader Node	
Importing Content from Multiple Files of the same Type to a Single Table	14
Importing Content from Multiple Sheets into a Single Table	
Saving an Excel File Excel Writer Node	
Adding a Sheet to an Excel File Excel Writer Node	
Updating Cells in an Existing Excel Sheet Excel Cell Updater Node	
Data Types in Excel Data Types in KNIME	
Connect to a Database Database Connector Nodes	
Tips on Reading Data with KNIME Analytics Platform	
Appending / Joining Data	
Appending Data Concatenate Node	
VLOOKUP Filter and Joiner Node	
Filtering and Transformations	
Filtering / Removing Rows with a Specific Value Row Filter	

Filtering / Removing Rows with Different Values Rule-based Row Filter	
Removing Duplicates Duplicate Row Filter	
Sorting Rows by Multiple Key Columns Sorter Node	
Removing Columns Column Filter Node	
Reordering and Renaming Columns Column Resorter and Column Rename Node	
Changing Data Types String to Number and Number to String	
Tip on Data Manipulation with KNIME Analytics Platform	
Data Aggregation	
Pivot Tables Pivoting Node	
Pivot Table without Columns GroupBy Node	
Unpivot Unpivoting Node	
Math Functions and Text Functions	
Math Functions Math Formula Node	
Math Functions Math Formula Node	
Concatenation and Find& Replace String Manipulation Node	
Formatting Excel Tables	
Creating an XLS Control Table with Tag Values	
Adding Formatting Actions based on Tag Values	

General Usage

Spreadsheets

Workflows and Nodes

Excel	KNIME Analytics Platform
Microsoft Excel is a spreadsheet program, which features calculation, graphing tools, pivot tables, and a macro programming language (Visual Basic for Applications, VBA for short). By using cell mathematics, macros, and VBAs you can edit a sheet. This can be really easy cell mathematics, like summarizing the values from cells A1 and B1 (= SUM(A1, B1)), but can be also really complex, embedded logic.	 KNIME Analytics Platform implements visual programming. This means that each data analysis step is represented by means of an icon block, called a node, in a graphical editor. Each node can perform one specific task. For example the Excern Reader node can read one sheet of an Excel file or the Row Filter node allows to filter rows based on a filter criterion. A sequence of connected nodes is called a workflow and is the corresponding concept of an Excel sheet with many functions and/or VBAs. Data is organized through data tables, where each data cell is identified by a column header and a Row ID. To visualize the content of a data table, see page 11.
A B C D E F G H J K L M N O P Q R S	Covering Filter Column Filter Excel Writer Image: Covering Filter Image: Covering Filter Image: Covering Filter Image: Covering Filter Image: Covering Filter Image: Covering Filter Image: Covering Filter Image: Covering Filter Image: Covering Filter Image: Covering Filter Image: C
	 For more details about KNIME Analytics Platform, check: S. Hayasaka, R. Silipo, <u>"KNIME Beginner's Luck"</u>, KNIME Press, 2021 R. Silipo, V. Palacios, <u>"KNIME Advanced Luck"</u>, KNIME Press, 2021

Folders

Excel	KNIME Analytics Platform
Excel files are normally saved in different folders. A single Excel file can contain multiple sheets.	The workspace defines the folder where all workflows, data and intermediate data are saved. One workflow corresponds to an Excel sheet with all formulas, visualizations and VBAs. All the projects and datasets saved in a workspace are available in the KNIME Explorer, located in the top left corner of the KNIME workbench. The path to the workspace is selected at the very beginning, after starting <i>KNIME Analytics Platform</i> .
	Cluber Select a directory as workspace KNIME Analytics Platform uses the workspace directory to store its preferences and development artifacts. Workspace: C:\Users\rosy\knime_3.5.0\workspace Browse Use this as the default and do not ask again Recent Workspaces OK Cancel
	You can still change the workspace after KNIME Analytics Platform has been launched, by going to "File" in the top menu and selecting "Switch Workspace". You can have multiple workspaces, e.g. for different projects or customers.
	Switch Workspace Other Preferences Export Preferences Import Preferences Install KNIME Extensions Update KNIME Exit

The KNIME Workbench

After downloading and installing KNIME Analytics Platform you can start it from the desktop or from the installation folder. The KNIME workbench, which you can see below, opens including the following panels:

"KNIME Explorer" showing the list of currently available workflows and KNIME servers for the selected workspace and the My-KNIME-Hub mountpoint.

"Workflow Coach" recommending the next node based on the KNIME user statistics and the node currently selected in the "Workflow Editor".

"Node Repository" containing all currently installed nodes. A "Search" box is available at top of this panel to search for nodes.

"Workflow Editor" in the center allowing for the creation and editing of workflows.

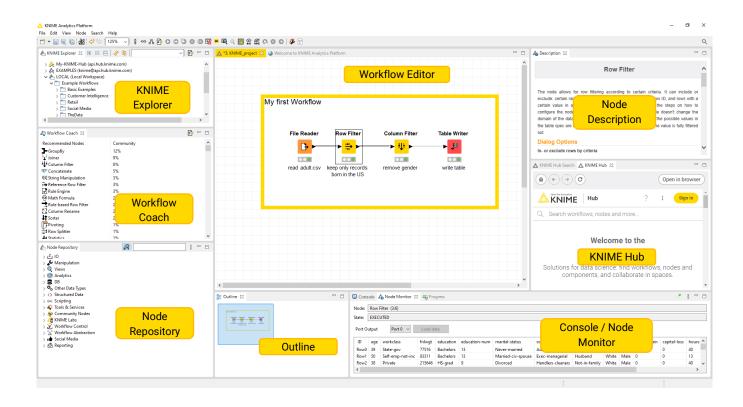
"Node Description" showing a text describing the node task and configuration settings, for the selected node either in the "Workflow Editor" or in the "Node Repository" panel.

"Node Monitor" showing a preview of the output table of the node selected in the "Workflow Editor"

"KNIME Hub" allowing use of the KNIME Hub to search for nodes, workflows, components, and extensions.

"Outline" offering an overview of the workflow

"Console" showing execution messages, e.g. error and warning messages.



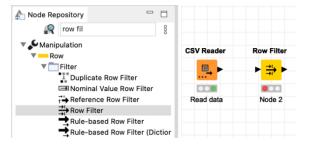
Building a KNIME Workflow

KNIME workflows are **created** by **dragging&dropping** nodes from the "Node Repository" or "Workflow Coach" panel to the "Workflow Editor". Use the search box on the top of the Node Repository or browse through the nodes, sorted by different categories to find the correct node for your next step.

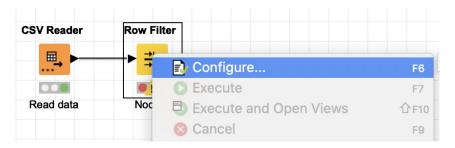
Nodes are **connected** to each other through their input and output ports. Just click the output port of the first node and release at the input port of the second node. Nodes that have just been created show a red light status: not yet configured. To **configure** a node, right-click the node and select the option "Configure" or alternatively double-click the node. The node "Configuration" window opens. Configure the node and close the configuration window. If the configuration is successful, the node status changes to a yellow traffic light.

The node is now configured, but not yet executed. To execute the node, right-click the node and select the "Execute" option. If the execution is successful, the node changes its status to a green light.

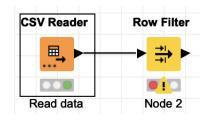




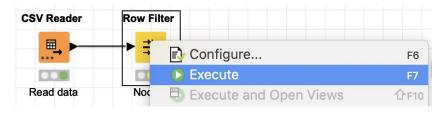
Step 3: Configure the node



Step 2: Connect the nodes



Step 4: Execute the node



Note 1: To create a new, empty workflow right click in the KNIME Explorer panel, select "Create New KNIME Workflow..." and define the name and destination of the new workflow in the new window.

Note 2: Click the magnifier next to the search box in the node repository to change the mode of the search box to a fuzzy search. This makes finding the correct node easier in the beginning.

Note 3: The "<u>Getting Started Guide</u>" guides you step by step through building your first example workflow.

Display Data Table

Excel	KNIME Analytics Platform
In Excel what you see is what you get. This means that the data table you see is the final data table.	The output data tables produced after node execution are always available. To see them: Right-click the node in the workflow Select the last option in the context menu Reset the last option in the context menu Reset the very Worldver designment of the workflow with the the second seco
	Column headerData TypeFile Hille Navigation ViewFile VariablesFile VariablesFile VariablesRow 0 39State-gow 7/250Sec - Columns: 14Poperties File VariablesRow 0 39State-gow 7/250Sec - Columns: 14Poperties File VariablesRow 0 39State-gow 7/250Sec - Columns: 14Poperties File VariablesRow 1 39State-gow 7/250Sec - maa.Weithing and Handlers-c., Husband Male 0 0Row 1 39State-gow 7/250Sec - maa.Weithing 100Row 2 38Private 284582Matters 14Matried-civ Exec-maa.Weithing 100Row 2 32Private 159449Bachelors 113Matried-civ Exec-maa.Withing 100Row 2 32Private 28687Matried-civ Exec-maa.Weithing 100Row 1 32

Input/Output

Opening an Excel File

Excel Reader Node

Excel	KNIME Analytics Platform
To open an Excel file you can either double click the file or open Excel, go to "File" n the top menu, select "Open" and then browse to the file you want to open.	The Excel Reader node reads a single sheet of an Excel file. Similar to Excel you first have to define the file path. Next you can select the sheet. The additional setting options enable you to define whether the table has column headers and / or row IDs and to specify which part of the sheet you want to read.
3 Reduce File Size 4 Share 5 Share 6 Restrict Permissions 7 Passwords 8 Page Setup 9 Print Area • Dockdod Drive • Dockdod • Dockd	File /Users/kathrinmelcher/Desktop/Product Data2.xls © Browse Sheet selection • • • Select first sheet with data (Product Data.xls_defa) • Sheet Select sheet with name Product Data.xls_defa • • Select sheet at index • • • Column header • • • • Table contains column names in row number 1 • • • Column headers • Column header • • • • Column headers • Contains column names in row number 1 • • • Column headers • Colerate row IDs • Table contains row IDs in column A • • Sheet area • Read entire data of the sheet • Read only data in columns from A • • and
Online Locations Options Cancel Open	rows from 1 to . (See "File Content" tab to identify columns and rows.) Preview File Content
	Preview with current settings The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab. Row 10 1 Custo (S) Products Row 11001 Private Investment Row 2 11001 Row 3 11001 Private Investment Row 3 11003 Row 4 11004 Private Investment Row 5 11005 Row 6 11004 Row 7 11005 Row 8 11004 Row 9 Cancel
	Note : Page 15 shows how to read and concatenate multiple Excel files which have the same column headers.

Opening a CSV or txt File

Excel

CSV Reader Node

	ose the file yo essary options				ing ch	aracto	er, to c	orrectly read the
	File Edit View Insert New New from Template	₩N ▼			Help			
Home Ins	Open Open Recent	HO ► A	Review	/iew ≫∙	📑 Ə Wrap 1	Text	General	
Paste 💞 Fo	Close Save Save As	₩W A - #S ?#S	= = =	◆ ≣ ◆ ≣	👥 Merge	& Center *	3 • %)
A	Save as Template Restore	>	E	F	G	н	I	
1	Import	-						
3 4	Red File Ole-							
5	Sha		Import					
6								
7	Resi Select the type	of file you wa	ant to impor					
8	Past							
9	Pag							
10		hat contain c	omma-sepa	rated value	es.			
11 12	Prin OHTML file							
13	Prop Hyper-text	markup files	from your o	omputer.				
14	Text file							
15								
16	Text files, v	which import	best when	separated	by tabs or	spaces.		
				Ca	ncel	Import		
			:: ≡ □□			≎		Q Search
				000 *				
		Favorites Dropbox (I All My File Field Desktop Application knime-woo	KNI Exc. s ve ns rksp	el Cheat Sheet (flowset.meta		i wine.csv wine.xlsx workflowse	t.meta	csv wine.csv
		Document	S					Cancel Get Data
								Cancer Get Data

To open a CSV file click "File" in the top menu and select "Import". Select CSV file

KNIME Analytics Platform

The <u>CSV Reader</u> node reads various text based files, e.g. CSV files. In the configuration window you can set all the necessary options, such as separating character, to correctly read the file.

	Dialog - 0:2 - CSV Reader	
Setting	Transformation Advanced Settings Limit Rows Encoding Flow Variables Memory Pol	File path
Input locati	on 🥂	
Read from	Local File System 🔹	
Mode	• File 🔿 Files in folder	
File	/Users/kathrinmelcher/Desktop/meter_data.csv 🗘 Browse	
Reader option	ins	
Autod	etect format	
6	olumn delimiter	
, ,	Delimiter	
	uote char " Quote escape char	
ų	duce char Quote escape char	
# C	omment char Column header	
	blumn header Has row ID	
Suppo	rt short data rows Prepend file index to row ID	
Preview		-
	gested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.	
	gested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.	
	gested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.	
The sug Row ID Row0	S Date S Time D Intensity 2007-01-01 00:01:00 10.4	
Row ID Row0 Row1	S Date S Time D Intensity 2007-01-01 00:01:00 10.4 2007-01-01 00:02:00 10.4	
The sug Row ID Row0 Row1 Row2	S Date S Time D Intensity 2007-01-01 00:01:00 10.4 2007-01-01 00:02:00 10.4 2007-01-01 00:03:00 10.4	
The sug Row ID Row0 Row1 Row2 Row3	S Date S Time D Intensity 2007-01-01 00:01:00 10.4 2007-01-01 00:02:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:04:00 10.4	
The sug Row ID Row0 Row1 Row2 Row3 Row4	S Date S Time D Intensity 2007-01-01 00:01:00 10.4 2007-01-01 00:00:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:05:00 10.4	
Row ID Row ID Row1 Row1 Row2 Row3 Row4 Row5	S Date S Time D Intensity 2007-01-01 00:01:00 10.4 2007-01-01 00:02:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:05:00 10.4 2007-01-01 00:05:00 10.4 2007-01-01 00:05:00 10.4	
The sug Row ID Row0 Row1 Row2 Row3 Row4	S Date S Time D Intensity 2007-01-01 00:01:00 10.4 2007-01-01 00:00:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:05:00 10.4	
Row ID Row ID Row1 Row1 Row2 Row3 Row4 Row5	S Date S Time D Intensity 2007-01-01 00:01:00 10.4 2007-01-01 00:02:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:05:00 10.4 2007-01-01 00:05:00 10.4 2007-01-01 00:05:00 10.4	
Row ID Row ID Row1 Row1 Row2 Row3 Row4 Row5	S Date S Time D Intensity 2007-01-01 00:01:00 10.4 2007-01-01 00:02:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:03:00 10.4 2007-01-01 00:05:00 10.4 2007-01-01 00:05:00 10.4 2007-01-01 00:05:00 10.4	

Note 1: Click the "Autodetect format" button if the node doesn't create the preview. **Note 2:** Check out the additional tabs to limit the number of rows or to change the encoding.

CSV Reader

...

Importing Content from Multiple Files of the same Type to a Single Table

Excel	KNIME Analytics Platform
If you have a folder containing multiple files of the same type (e.g. CSV), you can open them all at the same time in different Excel instances. Follow the instruction from the previous page and select all the files you want to open. To move data tables together into one single sheet, you have to proceed manually using copy and paste. Note: Before copying and pasting, ensure that all files have the same column order.	<text><text><image/></text></text>

Importing Content from Multiple Sheets into a Single Table

To move data from multiple Excel sheets into one Excel sheet you proceed manually, using copy and paste.	With a simple loop you can read all sheets of an Excel file automatically.
Note: Before copying and pasting, ensure that all files have the same column order.	The <u>Read Excel Sheet Names</u> node creates a list of all sheet names. The loop (the part in between the blue nodes) reads one sheet of the Excel file at each iteration. Therefore, at each iteration the <u>Table Row to Variable Loop Start</u> node creates a flow variable with the sheet name as its value. This flow variable is used in the <u>Excel Reader</u> node to control the sheet selection. The <u>Loop End</u> concatenates the content from the different tables. Read Excel Table Row To Sheet Names Table Row To Sheet Names Table Row To Sheet Names Table Row To Create table with Loop Start Excel Reader Loop End concatenates the all sheet names all sheet names Read one Combine all sheets into a single table iteration. Note 1: Lesson 3 of the free KNIME Self-Paced Course L2-DW KNIME Analytics Platform for Data Wrangles introduces flow variables.

Saving an Excel File

Excel Writer Node

Excel

To save the sheet of an Excel file you have different options:

- Click File->Save As... and define the output location
- Or press Shift+Ctrl+s and define the output location

KNIME Analytics Platform

The Excel Writer node writes or append the input data table into a sheet in an Excel file, in either xls or xlsx format. In the configuration window you can set the output location and sheet name. The additional setting options enable you to overwrite an existing file and to define whether you want to write the column headers / row ids into the first row / column of your Excel sheet.

File format & o	utput location	
Excel format	XLSX 📀	Output locati
Write to	Local File System 😌	
File	/Users/kathrinmelcher/Desktop/ExcelToKNIME.xlsx	S Browse
Write options	Create missing folders If exists: Overwrite • apper	nd 🔵 fail
Sheets		Sheet name
1. sheet name	default_1 🗘	oncername
Missing value I	nn headers column headers if sheet exists	Column headers and row key
Formulas		
Evaluate fo	rmulas (leave unchecked if uncertain; see node description for deta	ails)
Layout Autosize co	olumns Landscape A4 - 210x297 mm G	
-	ter execution	
open ne a	ICI EXECUTION	

Note 1: To write multiple tables into different sheets you can add dynamic input ports and define a sheet name for each input table.

Note 2: Select "append" for the settings "Write options" and "If sheet exists" to append the input data after the last row of an existing sheet.

Adding a Sheet to an Excel File

Excel Writer Node

Excel

To add a new sheet to an existing Excel file you have to click the plus below the table, next to the already existing sheets.

To add a table from another Excel file you can copy & paste the table manually.

AI Bit of JX Glasgow United Kin 17-171-83; Alois Berger 23.9.1972 47 Alois.Berge JX K 1 Glasgow United Kin 17-171-83; Alois Berger 23.9.1972 47 Alois.Berge K 2 Szczecin Poland 37-370-58(Michaela Schultz 9.6.1998 21. Michaela.5 (0 3 Sheffield United Kin 27-270-74; Rotraut Grünwald 20.4.1975 44 Rotraut.Gr (0 4 Bochum-H; Germany 64-647-95; Helga Heind 18.10.200(9 Helga.Hein (0	Idea	Editing	ells	Ce		ble v	nat as Tab Styles v		Number		Align	Font	lipboard	
Glasgow United Kin 17-171-83 Alois Berger 23.9.1972 47 Alois.Berge 0 2 Szczecin Poland 37-370-58(Michaela Schultz 9.6.1998 21 Michaela.S 0 3 Sheffield United Kin 27-270-74 Rotraut Grünwald 20.4.1975 44 Rotraut.Gr 0 4 Bochum-H: Germany 64-647-95; Helga Heindl 18.0.200 19 Heiga.Hein 0 0 5 Dortmund Germany 84-846-82; Mira Gleich 18.3.1997 22 Mira.Gleich 0 6 Valencia Spain 85-853-53; Joanna Radke 13.12.199 24 Joanna.Ref 7 Valencia Spain 65-655-25; Hanspeter Stort 25.1.1998 21 Hanspeter 1 8 Lodz Poland 96-969-844 Klaus-Peter Heinecke 13.4.1963 56 Klaus-Peter 9 9 Stuttgart Germany 48-484-59(Guenter Gasser 1.5.1988 <	L	V					6			-		v	L ,	A:
2 Szczecin Poland 37-370-58 (Michaela Schultz 9.6.1998 21 Michaela.5 Michaela.5 3 Sheffield United King 27-270-74 (Rotraut Grünwald 20.4.1975 44 Rotraut.GrÖ Michaela.5 Michael	L	ĸ	J	1	6							_		
3 Sheffield United Kin (27-270-74; Rotraut) Grünwald 20.4.1975 44 Rotraut.Gr M 4 Bochum-H: Germany 64-647-95; Helga Heindl 18.10.200 19 Helga.Hein M 1 5 Dortmund Germany 84-846-82; Mira Gleich 18.3.1997 22 Mira.Gleich M 1 6 Valencia Spain 58-582-55; Joanna Rake 13.1.199 24 Joanna.Rac 1 1 7 Valencia Spain 65-655-25; Hanspeter Storch 25.1.1998 21 Hanspeter 1 1 8 Lodz Poland 96-969-844 (Klaus-Peter Heinecke 13.4.1963 56 Klaus-Peter 1 1 9 Stuttgart Germany 48-484-59(Guenter Gassner 1.5.1988 31 Guenter.6 ² 1 0 Denver United Stat 47-474-554; Giuseppina Nitsch 9.12.1975 44 Giuseppina 1 1 1 Kharkiv Variane 8-585-44! Sami Zimmere 5.10.1982 37 <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>-</td> <td></td> <td></td> <td></td> <td></td> <td></td>									-					
4 Bochum-Hc Germany 64-647-95; Helga Heindl 18.10.200 19 Helga.Hein 10 5 Dortmund Germany 84-846-82; Mira Gleich 18.3.1997 22 Mira.Gleich 10 6 Valencia Spain 58-582-35; Joanna Radke 13.12.199; 24 Joanna.Rac 1 7 Valencia Spain 65-655-25; Hanspeter Storch 25.1.1998; 21 Hanspeter 1 8 Lodz Poland 96-96-944; Klaus-Peter Heineck 13.4.1963; 56 Klaus-Peter 0 9 Stuttgart Germany 48-484-59; Giuseppina Mira. 15.1988; 31 Guenter.Ge 1 0 Denver United Stat 47-474-55; Giuseppina Nitzch 9.12.1975; 44 Giuseppina 1 1 Kharkiv Ukraine 85-859-44; Sami Zimmerer 5.0.1982; 37 Sami Zimr1														
5 Dortmund Germany 84-846-82 Mira Gleich 18.3.1997 22 Mira.Gleich Mira.G					-									
5 Valencia Spain 58-582-35 Joanna Radke 13.12.199 24 Joanna.Rac1 7 Valencia Spain 65-655-25 Hanspeter Storch 25.11998 21 Hanspeter 1 8 Lodz Poland 96-969-844 Klaus-Peter Heinecke 13.4.1963 56 Klaus-Peter To 9 Stuttgart Germany 48-484-596 Guenter Gasner 1.5.1988 31 Gueren.Ge1 0 Denver United Stat 47-474-556 Giuseppina Nitsch 91.21975 44 Giuseppina 1 1 Kharkiv Ukraine 85-859-444 Sami Zimmerer 5.10.1982 37 Sami Zimr1					_									
8 Lodz Poland 96-969-84 (Klaus-Peter Heinecke 13.4.1963 56 Klaus-Peter Model 9 Stuttgart Germany 48-484-59 (Guenter Gassner 1.5.1988 31 Guenter, Ge ² 1 1 0 Denver United Stat 47-474-55 (Giuseppina Nitsch 9.12.1975 44 Giuseppina 1 1 1 Kharkiv Ukraine 85-859-44 (Sami Zimmerer 5.10.1982 37 Sami Zimm1 1			-											
9 Stuttgart Germany 48-484-59(Guenter Gassner 1.5.1988 31 Guenter.Ge1 0 Denver United Stat 47-474-555 Giuseppina Nitsch 9.12.1975 44 Giuseppina 1 1 Kharkiv Ukraine 85-859-444 Sami Zimmerer 5.10.1982 37 Sami Zimn1					.1	Hanspeter .	21	25.1.1998	Storch	Hanspeter	65-655-25	-	Valencia	7
Derver United Stat 47-474-551 Giuseppina Nitsch 9.12.1975 44 Giuseppina 1 1 Kharkiv Ukraine 85-859-441 Sami Zimmerer 5.10.1982 37 Sami Zimm 1					ro	Klaus-Peter 0	56	13.4.1963	Heinecke	Klaus-Peter	96-969-84	Poland	Lodz	В
1 Kharkiv Ukraine 85-859-44 Sami Zimmerer 5.10.1982 37 Sami Zimm 1					a 1	Guenter.Ga	31	1.5.1988	Gassner	Guenter	48-484-590	Germany	Stuttgart	9
					a 1	Giuseppina	44	9.12.1975	Nitsch	Giuseppina	47-474-55	United Stat	Denver	0
														1
2 Sofia Bulearia 13-130 BR Dana Wieczorek 30.3.1984 35 Ivana Wiec ⁶					0	Ivana Wiec 0	35	30.3.1984	Wieczorek			Bulgaria Custome	Sofia	2
· · · · · · · · · · · · · · · · · · ·	100%	- +	0		-						-			

KNIME Analytics Platform

The <u>Excel Writer</u> node can also add sheets to an existing Excel File. In the configuration window you can set the location of the existing Excel file and select "append" for the setting option "If exists". In the "Sheets" part you can define the new sheet name and whether the node should overwrite or fail in case a sheet with the defined sheet name exists already.

•	Dialog - 3:99 - Excel Writer	
	Settings Flow Variables	Location of the existing File
File format & output location		
Excel format XLSX 😒		
Write to Local File System	n 😌	
File /Users/kathrinme	elcher/Desktop/ExcelToKNIME.xlsx	Srowse
Write options 🗌 Create missing	g folders If exists: 🔿 overwrite 🔾 app	pend 🔿 fail
Sheets		
1. sheet name default_1	•	Append option
If sheet exists O overwrite O	append 🧿 fail	
Names and IDs Write row key		Sheet settings
 ✓ Write column headers ✓ Don't write column headers if 	f sheet exists	
Missing value handling		
Replace missing values by		
Formulas		
Evaluate formulas (leave unch	necked if uncertain; see node description for d	letails)
Layout		
Lujout		
Autosize columns		
Autosize columns	A4 - 210x297 mm 😒	
Autosize columns	A4 - 210x297 mm 🕄	
 Autosize columns Portrait Landscape 	A4 - 210x297 mm 😮	Apply Cancel ?

Excel Writer

чx

Updating Cells in an Existing Excel Sheet

Excel Cell Updater Node

KNIME Analytics Platform
The Excel Cell Updater node updates the cells in an existing Excel sheet, based on an input data table. The input table needs a column with the cell address that should be updated e.g., B2 or 2:2. In addition the table needs one column for each datatype with the new cell content. Each row is only allowed to have one value.
 Dialog - 4:2 - Excel Cell Updater (Update Values) Settings Flow Variables Job Manager Selection
Input file
Read from Local File System 😳
File /Users/kathrinmelcher/Desktop/old-file.xlsx 🗘 Browse
Output file
Create new file
Write to Local File System 😋
File /Users/kathrinmelcher/Desktop/new-file.xlsx Image: Browse Write options Create missing folders If exists: • overwrite fail
Update
1. Excel sheet Sheet1 Sheet1 Based on address column SAddress
Missing value handling Address column Address values by
Formulas
Evaluate formulas
OK Apply Cancel
Note 1 : Activate the checkbox "Evaluate formulas" to evaluate all formulas using the new cell content.
Note 2 : Activate the checkbox "Create new file" to keep the original file.
Note 3: Example Workflow on the KNIME Hub.

Data Types in Excel

Data Types in KNIME

Excel		KNIME Analytics Platform	
The screenshot on the right shows y different available datatypes in Exce They can be mapped to the followin types in KNIME Analytics Platform	ABC General 123 No specific format 123 Number Currency Accounting C Short Date C Long Date C Time % Percentage 1/2 Fraction 10 ² Scientific ABC Text More Number Formats	Transformations	rtics Platform tries to autodete e this during reading you can us node. Excel Reader ettings Flow Variables Memory Policy C Enforce types Take columns from: Un
Excel	KNIME Analytics Platform		
General	General Number or String		
Number	Number	A 7	•
Currency	Number (Integer or double)		Preview File Content
(e.g. 50,25 €)	(e.g. 50,25)	Preview with current settings The suggested column types are based on t	he first 10000 rows only. See 'Advanced Settings
Accounting	Number (Integer or double)		abar D. Curra D. Accou, D. Short D. D. Long

Accounting Number (Integer or double) (e.g. 50,25 €) (e.g. 50,25) Short Date Local Date Long Date Local Date Local Time Time Number (double) Percentage (e.g. 50%) (e.g. 0,5) Number (double) Fraction Scientific Integer (e.g. 5,00E+02) (e.g. 500)

 S Gen Num D Curr D Accc S Shor Long C Time D Perc S Fracc I Sciel I Text 	ency unting t Date Date e entage tion						string Number (integer) Number (double) Number (double) Local Date Local Date
D Curr D Acco Shor D Long D Time D Perc S Frac L Scient	ency unting t Date Date e entage tion				[] [] [] []		Number (double) Number (double) Local Date
D Acco S Shor D Long D Time D Perc S Frac	unting t Date Date entage				6		Number (double)
Image: Shore Image: Shore	t Date Date entage				5	II L	local Date
 J Long J Time D Perc S Frac I Scient 	Date entage				5	_	
D Percos S Frac	entage tion				-	۱L	ocal Data
D Perco	entage tion						Local Date
S Frac	tion				6	ЭL	local Time
Scier					[DN	Number (double)
_					\$	S S	String
Text	itific					I N	Number (integer)
						I N	Number (integer)
? <an< td=""><td>y unknown r</td><td>new colum</td><td>n></td><td></td><td>Ľ</td><td>?</td><td></td></an<>	y unknown r	new colum	n>		Ľ	?	
					0		
				Prev	iew Fi	ile (Content
ne sugges Row ID	ted colum	n types ar		Number	D Curre.		
			ОК		Apply		Cancel
	ew with cu	ew with current setti he suggested column Row ID [1] Gener.	ew with current settings he suggested column types an Row ID 1] Gener [5] Gen	ew with current settings he suggested column types are based of Row ID I Gener S Gener I w0 3 Hallo 4	Prev ew with current settings he suggested column types are based on the fir Row ID 11 Gener [\$] Gener [1] Number W0 3 Hallo 4 4	Preview Fi ew with current settings Fi he suggested column types are based on the first 10000 Row ID Row ID I Gener S Gener I Number W0 3 Hallo 4 4.5	ew with current settings the suggested column types are based on the first 10000 ro Row ID I Gener S Gener I Number D Curre w0 3 Hallo 4 4.5

Note 1: The Transformation tab is also available in other reader nodes. Note 2: In the Transformation tab you can also rename, remove, and change the order of columns.

Note 3: To change the data type later in the workflow you can use one of the following nodes: String to Number, Number to String, or Table Manipulator.

Connect to a Database

Excel

databases/

Database Connector Nodes

KNIME Analytics Platform You can connect to a database in Excel. A description about how to do this is Reading data from a database follows three steps: connect, select, and extract. The available via the following link: workflow is built step by step with a Connector, a DB Table Selector, and a DB http://www.erpsoftwareblog.com/2017/01/microsoft-excel-connections-sql-Reader node. Database nodes simply build the SQL query, they do not execute it. Only the final node, e.g. the <u>DB Reader</u> node, executes the SQL query and extracts the data. DB Reader SQLite Connector **DB Table Selector** Ē. 25 Connect to Database Select a Table Execute SQL query Microsoft SQL Microsoft Access A number of database **DB Connector** H2 Connector Server Connector Connector connector nodes are ġ, Вн 🗖 BA available to connect to the most commonly used databases. However, the MySQL Connector Vertica Connector Oracle Connector **DB** Connector node allows Ēм Ev 🛛 vou to connect to all JDBC 20 compliant databases. There are more database nodes to help build a SQL query for in database processing. You can use them in between the <u>DB Table Selector</u> and the <u>DB Reader</u>

node. **DB Column Filter DB Table Selector DB Reader** - ## 2→ => **DB Row Filter** DB Reader DB Table Selector **DB GroupBy DB** Joiner - E - 13 -

Tips on Reading Data with KNIME Analytics Platform

All reader nodes require a path to the input file location. Let's collect some Tips&Tricks for this:

Tip&Trick 1: Use drag&drop from the KNIME Explorer:

Data files saved in the workspace folder are available in the KNIME Explorer panel (top left panel). To read in one of these files, you just drag&drop the file from the KNIME Explorer panel to the workflow editor. KNIME automatically creates the correct reader node and sets the path of the input location.

Tip&Trick 2: Different options to define a file path:

In KNIME we have different options to provide a file path. This becomes important when you start sharing your workflows or exporting them to other KNIME Analytics Platform installations or KNIME Servers. There are four default file systems available in KNIME Analytics Platform.

- Local File System: Allows you to select a file/folder from your local system.
- Mountpoint: You can connect to a KNIME Server or the KNIME Hub via additional mountpoints in the KNIME Explorer. To read data from either LOCAL or another mountpoint select "Mountpoint". When selected, a new drop-down menu appears so that you can choose the mountpoint. Unconnected mountpoints are grayed out but can still be selected (note that browsing is disabled in this case). Go to the KNIME Explorer and connect to the mountpoint to enable browsing.
- *Relative to:* Allows you to choose whether to resolve the path relative to the current mountpoint, current workflow, or the current workflow's data area. When selected a new drop-down menu appears to choose which of the three options to use.
- *Custom/KNIME URL*: Allows to specify a URL (e.g. file://, http:// or knime:// protocol). Browsing is disabled for this option.

Tip&Trick 3: Reading from another file system:

KNIME Analytics Platform allows you to connect and read from many different sources / file system, e.g. Amazon S3, Microsoft SharePoint Online, Databricks to name just a few. Three steps are necessary (the <u>file handling guide</u> gives you further information).

Step 1: Click "..." in the bottom left corner of the reader node icon to add a *File System Connection port*



Step 2: Connect to the desired file system via the dedicated connector node and connect it with the reader node



Step 3: Select the file/folder in the connected file system

-Input locati	on	
Read from	Amazon S3	~
Mode	● File ○ Files in fo	older
File		

Appending / Joining Data

Appending Data

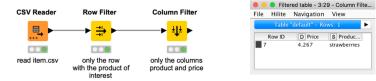
Concatenate Node

Excel	KNIME Analytics Platform	
To append data to a table select manually the area you want to append and copy and paste the content below the first table. Note 1: Before copying and pasting ensure that all tables have the same column structure.	Concatenate	add an additional input port, click on the e and select "Add input port". Concatenate

VLOOKUP

KNIME Analytics Platform Excel The VLOOKUP function is used for different tasks. The most common uses are: 1. To look up a certain value, e.g. the price of a certain product 😑 😑 Filtered table - 3:29 - Column Filte... CSV Reader Row Filter **Column Filter** File Hilite Navigation View 1. Look up a certain value, e.g. the price of a certain product. <u>⊟</u> ▶<mark>-1</mark>1+ ▶ D Price S Produc.. Row ID 4.267 strawberries 2. Join columns based on a primary key (look up value), e.g. join product information based on the product ID.

An alternative function for the second task is INDEX MATCH.



Note 1: Your full original table is still available at the output port of the Table Reader node. See more information about the Row Filter and Column Filter nodes on pages 26 and 30.

2. Join columns based on a joining column, e.g. join product information based on product ID.

Match 🔾 all of the Top Inpu	followingany of the any of	he following Botto	m Input ('right' table)		Join colu	(0)	
S Promotion					+ -		
					+		
Compare values in j	oin columns by 💿 val	ue and type 🔵 stri	ng representation 🔵 r	naking integer t	ypes compatible		Join
Include in output						_	
Matching rows			Left outer join		Join mode		1.
🗹 Left unmatched i	ows					—	_▶ ◀
Right unmatched	TOWE						
Right unmatched	Tows			J			
Output options							
Route unmatchee Merge join colun	f rows to separate ports	5					
Hiliting enabled	ns						
Row Keys							
	inal row keys with sepa	arator _					
Assign new row							
			OK Apply	Cancel	0		

Note 2: For task 2. you can also use the Cell Replacer node

Filtering and Transformations

Row Filter

Excel		KNIME Analytics Platform
To filter data select a random cell in the data table, go to the Home tab, click the "Sort& Filter" button and select "Filter".	A Sort & Filter A Filter A Custom Sort Filter Custom Sort Filter Reapply	In KNIME Analytics Platform there is no difference between filtering and removing rows, as the original table is not deleted and is still available at the output port of the previous node.
		Filter Criteria Flow Variables Memory Policy Filter column
Select the value you are interested in from	E F G H I Area Code V State V Phone V	Column value matching
the drop down menu.	Here Sort 415 Sort 415 \$2 + Ascending 418 \$2 + Ascending 408 By color: 510 By color: 510 Choose One 415 \$4 + Seending 416 \$10 + 1000000000000000000000000000000000	Column to test: S ProductName Column to test: ProductName
To remove rows, select the rows you want to o Rows".	delete, right click and select "Delete	OK Apply Cancel
		 Note 1: On the left you can choose whether you want to include or exclude the rows with the matching value Note 2: If you only interested in the rows with one specific value you can use the Row Filter node. Note 3: If you want to include rows based on different values you can use the Rule-based Row Filter. (See next page). Note 4: Further filter options are available, e.g. on a numerical range, filter rows by row number or row ID, or missing values only.

Filtering / Removing Rows with Different Values

Rule-based Row Filter

n KNIME Analytics Platform there is no difference between Rule-based Row Filter Iltering and removing rows, as the original table is not deleted
The <u>Rule-based Row Filter</u> node filters rows in or out according to a set of rules. To include / exclude rows with two different values you can use the following expression Column_Name\$ = "Value 1" OR \$Column_Name\$ = "Value 2"=> TRUE
Column List ROWD ROWODXX Description ROWODXX All Checks whether the value of the left expression is like the wildcard pattern defined by the right expression VMail Message Day Mins Function Checks whether the value of the left expression is like the wildcard pattern defined by the right expression Di VMail Message Day Mins ? < ? ? I VMail Message Day Charge ? < ? ? Day Calls ? > ? ? Day Calls ? > ? ? Day Calls ? > ? ? Day Calls ? < ? ? Pow Carlable List © Knime.workspace Power and the column names > 5.9 ⇔ FALSE Expression 1 // enter ordered set of rules, e.g.: ? // stouble column names Like "⇔blues" ⇒ FALSE 4 // TRUE ⇒ TRUE 5 \$Area Code\$=40B OR \$Area Code\$= 415 ⇒ TRUE * FALSE * TRUE
b

Note 1: At the bottom of the configuration window you can choose whether you want to include or exclude TRUE matches.

OK Apply Cancel 🕐

Note 2: Columns are given by their name surrounded by \$. Add them to the expression frame by double clicking a column name in the Column List. **Note 3**: The <u>Rule-based Row Filter</u> node has a number of different functions for many advanced filter options.

Note 4: Different rows in the expression frame work like an OR conjunction.

Removing Duplicates

Duplicate Row Filter

Excel	KNIME Analytics Platform
 o remove duplicates Select the range of cells that might have duplicates that you want to remove. Go to the Data tab and select "Remove Duplicates" Data Review View Q Tell me	The <u>Duplicate Row Filter</u> node detects and treats duplicates. The default treatment removes duplicate rows like in Excel. The columns in the "Include" frame correspond to the selected columns for duplicate detection in Excel. This means rows that have the same values in these columns are detected as duplicates.
Â↓ ZAZ Clear Clear	Dialog - 0:35 - Duplicate Row Filter
Geography Z A Sort Filter Advanced Advanced Columns Remove Duplicates	Options Advanced Flow Variables Memory Policy
	Choose columns for duplicates detection
Select the columns for duplicate detection and click "OK"	Manual Selection
	T Filter
Remove Duplicates	S City S FirstName
✓ My list has headers	S Country D Age
Select All	S Email
Column A	٢.
Column B	>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
Column C	
Column D Column E	Enforce exclusion Enforce inclusion
Column F	
Cancel	OK - Execute Apply Cancel
	Note 1 : In the "Advanced" tab you can change the treatment for duplicates, for
	example to keep duplicate rows and to add a column showing which of the row
	are unique, chosen, or duplicates.

Sorting Rows by Multiple Key Columns

Sorter Node

Excel	KNIME Analytics Platform
To sort rows by multiple columns, select the columns by which you Then click on Sort&Filter and select "Custom Sort".	vant to sort. With the <u>Sorter</u> node you can sort by one or more columns in either ascending or descending order.
Sort	
Add levels to sort by:	Dialog - 0:318 - Sorter
Column Sort On Order Color/Icon Sort by Day Mins \$ Values \$ Largest to Small \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$	Sorting Filter Advanced Settings Flow Variables Memory Policy
Then by Night Mins 🗘 Values 🗘 Largest to Small 🗘 🗘	
Then by Eve Mins O Values O Largest to Small O O	
	Sort by
	Descending
+ - Copy Options Cancel OK	
	↑↓ i
	CallActivity O Ascending
	Descending
	+ Add Rule
	OK Apply Cancel
	Note 1 : You can add as many key-columns as you want by clicking the "Add Ru button.
	Note 2: You can temporarily sort the output table of a node. Click on the colum
	header based on which you want to sort and select whether you want to sort ascending or descending.

Removing Columns

Column Filter Node

Excel	KNIME Analytics Platform
To remove columns just select and delete the superfluous columns. Remember: If you remove a column used in a calculation field this will break your formula.	With the <u>Column Filter</u> node you can delete superfluous columns. Remember that the full data table will be still available at the output port of the previous node and for calculations you run upstream.
	Dialog - 6:7 - Column Filter
	Column Filter Flow Variables Memory Policy
	Manual Selection Wildcard/Regex Selection Type Selection Exclude
	TFilter
	I CustServ Calls I VMail Message I Day Calls D Day Mins B Eve Mins Intl Mins I not Mins Intl Mins I Day Charge Eve Calls I Eve Calls Eve Charge I Night Calls Night Charge
	Enforce exclusion Enforce inclusion
	OK Apply Cancel
	 Note 1: You can use the arrow buttons in the middle to move columns from the Include to the Exclude frame and vice versa. Note 2: You can use the Wildcard/Regex Selection to automatically remove columns by a name patter. Note 3: You can use the Type Selection to automatically remove columns by a type.

Reordering and Renaming Columns

Excel

٠

٠

٠

٠

To reorder columns:

Column Resorter and Column Rename Node

KNIME Analytics Platform To resort columns you can use the Column Resorter node. Define the desired Select the column header of the column you want to move order of the columns by selecting the one you want to move and using the "Actions" buttons on the right. Press ctrl + shift + down arrow Click and hold the green outline of the column you want to move Dialog - 3:19 - Column Resorter Drag your column to the desired position Resort columns Flow Variables Memory Policy Columns D fixed acidity Actions D volatile acidity A-Z D citric acid **Column Resorter** D residual sugar Z-A D free sulfur dioxide D chlorides Up D total sulfur dioxide D density Down D pH D sulphates Move First D alcohol Move Last I quality ? <any unknown new column> Reset OK Apply Cancel

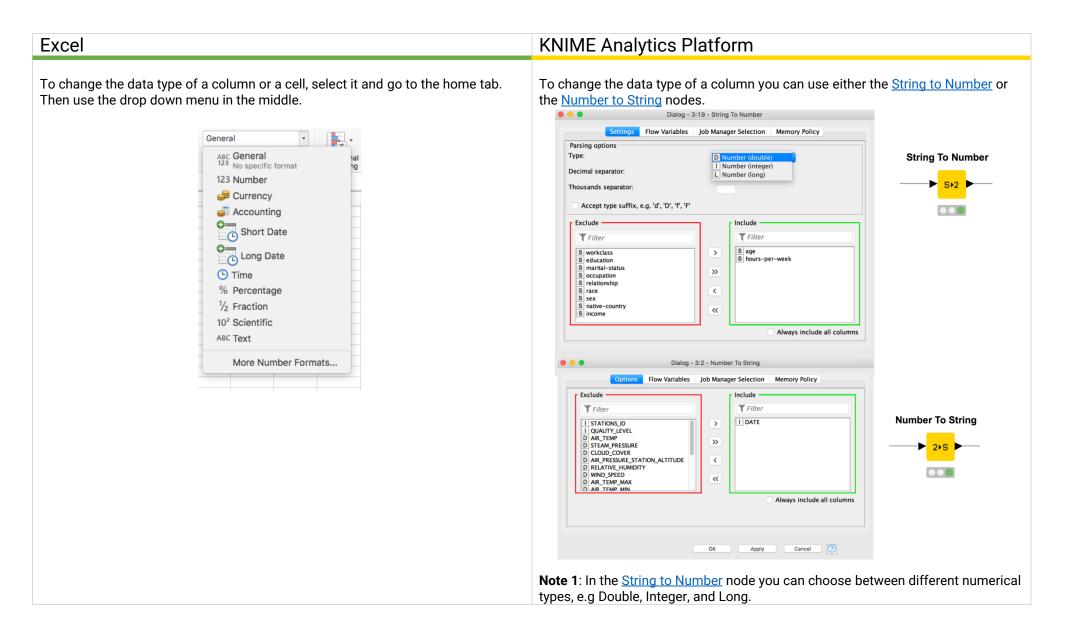
To rename a column just click on the column cell and change the cell value.

To rename columns you can use the Column Rename node. Double click the column you want to rename, activate the checkbox "Change" and define the column header in the textbox.

	•••	Dialog - 3:22 - Column Rename	
		Change columns Flow Variables Memory Policy	
	Column Search	quality	Remove
Column Rename	Filter Options None	Change: Target	ue
── <mark>₽</mark> ₽	D fixed acidity D volatile acidity D citric acid		
	D residual sugar D chlorides D free sulfur dioxide D total sulfur dioxide		
	D density D pH D sulphates D alcohol		
	quality	•	
		OK Apply C	ancel

Changing Data Types

String to Number and Number to String



Tip on Data Manipulation with KNIME Analytics Platform

Table Manipulator



The <u>Table Manipulator</u> node allows you to perform a lot of the described transformations in one node. You can use it to:

- Concatenate multiple tables (after adding dynamic input ports by clicking on ... on the lower left of the node)
- Filter, resort, and rename columns
- Change the data type of column

•			Dialog - 0:3 - Table Manipulator	
			Settings Flow Variables Memory Po	licy
_		D Prepe	Resort columns	
nove	e co	olumns		
R	eset	ons 🔶 Move up	↓ Move down ☑ Enforce types Take co	olumns from: 💿 Union 🗌 Intersection
	┙	Column	New name	Туре
		1 age		I Number (integer)
н		S workclass	New Name	S String
		1 fnlwgt		Number (integer)
		S education		S String
		l education-num	Change column name	
		S marital-status		S String
	—	S occupation		
	_	S relationship		S String Change data type
		S race		S String
	—	S sex		S String
				Number (integer)
		l capital-gain		
		capital-loss		Number (integer)
		l hours-per-week		Number (integer)
		S native-country		S String
	<	S income		S String
		? <any colur<="" new="" td="" unknown=""><td>nn></td><td>?</td></any>	nn>	?
A 7	-		0	
Prev				
O [Data	analysis successfully comp	pleted.	
	Davis			
R	Row ow0	ID I age S New 39 State-go		marital-s S occupa S relation ever-married Adm-clerical Not-in-family
	ow1			arried-civ Exec-mana Husband
R	ow2	38 Private	215646 HS-grad 9 Di	ivorced Handlers-c Not-in-family
	ow3	53 Private		arried-civ Handlers-c Husband
R	ow4	28 Private	338409 Bachelors 13 M	arried-civ Prof-speci Wife

Data Aggregation

Pivot Tables

Pivoting Node

Excel				KNIME Analytics Platform	
 To create a pivot table in Excel Click on any single cell inside the Go to the Insert tab and click Pivo Select table / range and output 	ot Table	Create PivotTable		 The <u>Pivoting</u> node is configured via three tabs: "Groups" defines the group columns, aka columns in the "Rows"-frame in Excel (final row IDs) "Pivots" defines the pivoting columns, aka columns in the "Columns"-frame in Excel (final column headers) "Manual Aggregation" corresponds to the "Value" setting option. Select one or more columns for aggregation from the available columns list and select an aggregation method for each selected column. 	
location	Choose the data that you want to analyze. Select a table or range Table/Range: CallsDatal\$A\$1:\$P\$3334 Use an external data source			The <u>Pivoting</u> node produces three output tables: the pivot table and the total values for columns and rows.	
	Choose conn Choose where to place New worksheet Existing workshe Table/Range:	e the PivotTable.	Is have been retrieved.	Dialog - 3:16 - Pivoting Description Flow Va Values Column Policy Values Groups Pivots Manual Aggregation Aggregation settings Available columns Select To change multiple columns use right mouse click for context menu. I age Mean S workclass Sincome First Column	
 Choose pivot table fields by dragg into "Columns", "Rows" and "Value 		PivotTable Field Formula Bar FIELD NAME Oay Mins Eve Mins Night Mins	Q Search fields	Image Concatenate I education-num add >> Smartial-status First Socupation add all >> Srelationship add all >> I capital-gain << remove	
		♀ Filters	III Columns	Advanced settings Column naming: Aggregation method (column name) C Enable hiliting Process in memory Retain row order Maximum unique values per group 10,000 C Value delimiter ,	
		E Rows	Σ Values	OK Apply Cancel	
		Drag field	Is between areas	Note 1 : The <u>Pivoting</u> node doesn't have "Filter" options, but you can simply use a <u>Row Filter</u> node beforehand.	
				Note 2 : In KNIME you have to choose at least one column for the Groups and Pivots. In case you want to choose only "Rows" you can use the <u>GroupBy</u> node.	

Pivot Table without Columns

GroupBy Node

Excel	KNIME Analytics Platform
Create a pivot table as described on the previous page and drag only columns into the "Rows" and "Values" frames.	 The <u>GroupBy</u> node is configured via two tabs: "Groups" defines the group columns, aka columns in the "Rows" frame in Excel (final row IDs) "Manual Aggregation" corresponds to the "Value" setting option. Select one or more columns for aggregation from the available columns list and select an aggregation method for each selected column. The "Value" setting option corresponds to the "Manual Aggregation" tab.
	Dialog - 4:15 - GroupBy
	Rows Settings Description Flow Variables Job W Values Memory Policy
	Groups Manual Aggregation Pattern Based Aggregation Type Based Aggregation Group settings
	Available column(s) Group column(s)
	Filter 1 age § workclass 1 fnlwgt education-num § marital-status § occupation § relationship § race 1 capital-loss 1 hours-per-week § native-country § income
	Advanced settings Column naming: Aggregation method (column name) Maximum unique values per group 10,000 Value delimiter .
	OK Apply Cancel

Unpivot

Note 1: The unpivot command is available without any additional downloads in Excel 2016 for Windows. If you are using a different version, you may need to first

download the free Power Query add-in from the Microsoft site. Authoring in the

Power Query Editor is not supported for Mac yet.

Unpivoting Node

 Store the dataset in a table. Select any cell in the table. Click the "Data" tab and select "From Table/Range". 	 The <u>Unpivoting</u> node is configured via one tab: In the upper section, "Value columns" defines the column(s) to unpivot, aka column selection in Excel. In the lower section, "Retained columns" defines the column(s) that remains unchanged, aka unselected
 Get From From Table/ Recent Existing Sources Connections All ~ Befresh All ~ Befresh Connections All ~ Befresh Connections All ~ Cueris This opens the "Power Query Editor". Select the columns to unpivot by holding down the shift key. Click the "Transform" tab of the Power Query Editor and select "Unpivot Columns". I able 1 - Power Query Editor Table 1 - Power Query Editor Tansform Add Column View Columns To Columns To Columns To Columns Columns 	Columns in Excel. Options Flow Variables Manual Selection Wildcard/Regex Selection Exclude include Filter Sidoperations Options Skip rows containing missing cells Retained columns Skip rows containing missing cells
By as Headers → 🔂 Count Rows 🖳 Rename 🌳 Prot Column 📋 Convert to List Column → Convert to List Column → Text	Manual Selection Wildcard/Regex Selection Type Selection Exclude Include Filter Exclude Filter Exclude Filter Filter
	Exclude Include Y Filter Filter S Marketing S Account S Operations Account
Table Any Column Text f_x = Table.TransformColumnTypes(Source, {{"Account", type text}, {"Marketing", Int64.Typ f_x = Table.TransformColumnTypes(Source, {{"Account", type text}, {"Marketing", Int64.Typ f_x = Table.TransformColumnTypes(Source, {{"Account", type text}, {"Marketing", Int64.Typ Im. $A_{C_c}^{R_c}$ Account 1^2_3 Marketing Im. $A_{C_c}^{R_c}$ Account 1^2_3 Marketing	Filter Filter S Marketing S Account
TableAny ColumnTextTableTable.TransformColumnTypes(Source, {{"Account", type text}, {"Marketing", Int64.Type a_{B_c} <td>Exclude Include T Filter Filter S Marketing S Account S Comporate S Account</td>	Exclude Include T Filter Filter S Marketing S Account S Comporate S Account
Table Any Column Text > f_x = Table.TransformColumnTypes(Source, {{"Account", type text}, {"Marketing", Int64.Typ ::::::::::::::::::::::::::::::::::::	Exclude Filter S Marketing S Operations S Corporate Include Filter S Account S Account

Note 1: the <u>Unpivoting</u> node requires selecting at least one column in the "Value columns" section.

Math Functions and Text Functions

Math Functions

Math Formula Node

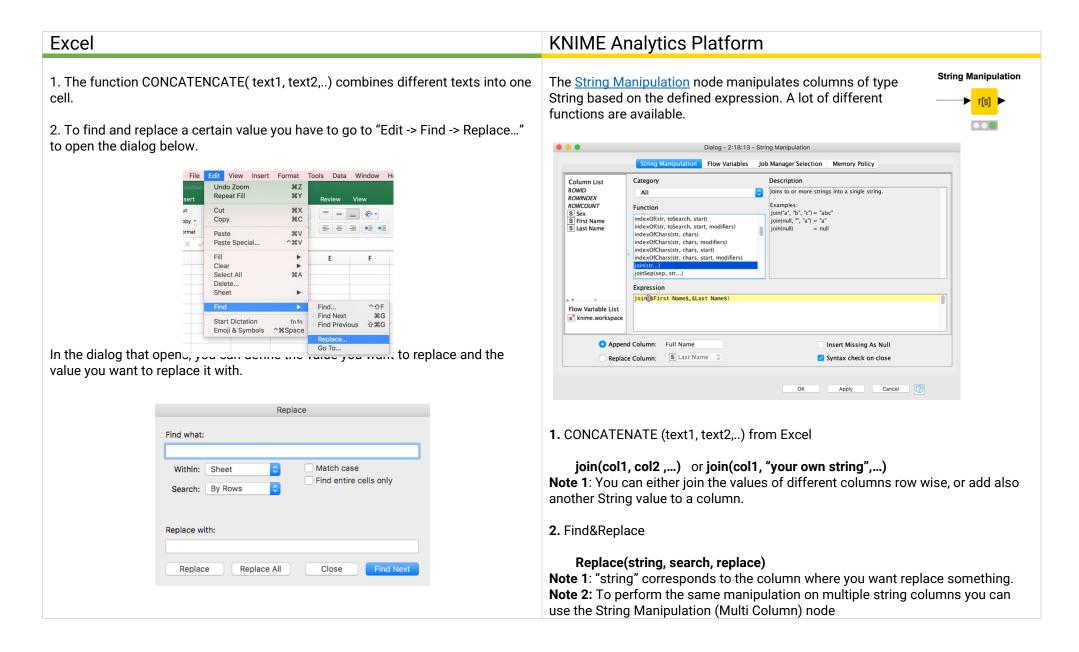
Excel	KNIME Analytics Platform
Here is a list of frequently used math formulas in Excel. You will find a translation into the KNIME formulas on the next page: • SUM • AVERAGE • MEDIAN • SUMPRODUCT • ABS • SUMIF • Round functions • ROUNDUP • ROUNDDOWN If you're often used math formula is missing sent a message to kathrin.melcher@knime.com.	The Math Formula node evaluates a mathematical expression based on the values in a row. It has a long list of functions. The table on the next page provides a translation of the most used Excel functions.

Math Functions

Excel	KNIME Analytics Platform
SUM(number1;number2;)	COL_SUM(Col)= Sum of the values in the selected column.Col1 + Col2.= Sum of the values in Col1 and the values in Col2 for each row.
AVERAGE(number1;number2;)	COL_MEAN(col_name) = Average of the values in the selected column. average(Col1, Col2,) = Average of the values of the selected columns for each row.
MEDIAN(number1;number2;)	COL_MEDIAN(Col_name) = Median of the values in the selected column. Median(Col1, Col2,) = Median of the values of the selected columns for each row.
SUMPRODUCT(number1;number2;)	Sequence of two "Math Formula" nodes: First one: Multiply the two columns using the expression Col1*Col2 and Append a new column. Second one: Use COL_SUM on the new column.
ABS(number1)	abs(Col) = The absolute value for all values in the selected column
SUMIF(range; criteria)	Sequence of two "Math Formula" nodes: First one: Append new column with if(criteria, column to sum ,0) Second one: Use COL_SUM on the new column
Round functions: ROUND(number1, num_digits) ROUNDUP(number1, num_digits) ROUNDDOWN(number1, num_digits)	round(Col, NumberOf Digits) = Number of digits is optional. ceil(Col*10^(num_digits))/10^num_digits floor(Col*10^(num_digits))/10^num_digits

Concatenation and Find& Replace

String Manipulation Node



Formatting Excel Tables

In chapter 1 we introduced the Excel Writer, which you can use to write your result table into an Excel Sheet. By default, this is a simple table without any formatting like colors, border cells, etc. In this chapter, we want to show you how to use the XLS Formatting nodes of the community extension <u>Continental Nodes</u> <u>for KNIME</u>. These nodes enable you to add formatting instructions and advanced settings to already existing XLS files, so that you can create Excel reports that have the look and feel you used to.

Standard written table

Home K4	insert	Draw Page I	ayout Formulas Data	Review View	1	Share Comment
-	A	В	С	D	E	F
1	Year	Quarter	Store - no CC	Store - with CC	OnlineStore	
2	2019	1	36862,74	66775,81	114196,84	
3	2019	2	38059,65	70483,79	113399,81	
4	2019	3	48149,06	76791,58	96116,79	
5	2019	4	47220,13	61563,41	105625,31	
6						
7						
4	default	+			T B P	

Styled table

Home A1	insert		Layout Formulas Data	Review View		Share Comment
/	A	В	С	D	E	F
1	Year	Quarter	Store - no CC	Store - with CC	OnlineStore	
2	2019	1	36862,74	66775,81	114196,84	4
3	2019	2	38059,65	70483,79	113399,8	1
4	2019	3	48149,06	76791,58	96116,7	9
5	2019	4	47220,13	61563,41	105625,3	1
6						
7						

Continental Nodes for KNIME

XLS Formatter Nodes

Figure 1 On the left you can see an Excel table created by an <u>Excel Writer</u> node without formatting, and on the right a styled table after formatting information have been added with the XLS Formatter nodes, e.g. yellow background for the headers.

The key to your formatted Excel sheet is an additional XLS Control Table of the same size as the original file with one or more comma separated tag values, e.g. header, border, etc. Different XLS Formatter nodes assign different formatting instructions to the cells based on these tags, e.g. you can change the background color for all cells that are tagged "header". Your formatting can then be applied to an already existing Excel sheet with the <u>XLS Formatter (apply)</u> node.

Hint: Use a flow variable connection to make sure that the Excel files are already written.

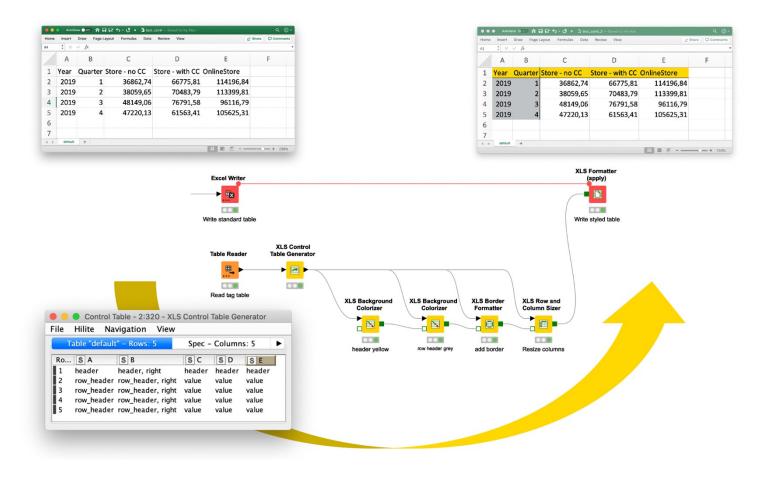


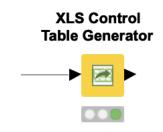
Figure 2 Bottom left you can see a control table with tag values, which is the key for your styled table. Based on the tag values the yellow XLS formatting nodes collect formatting instructions, which are them applied by the <u>XLS Formatter (apply)</u> node, producing the styled table (top right).

To summarize, this means we have to perform the following steps:

- 1. Write the table to an Excel Sheet
- 2. Create an XLS Control Table with tag values
- 3. Add formatting instructions based on tag values
- 4. Apply the formatting instructions to the existing Excel Sheet

This chapter is divided into two sections. The first section of this chapter shows two ways of creating an XLS Control Table with tags. The second section introduces some of the nodes that are available to add formatting instructions.

Hint: You can't find the nodes in your node repository? The <u>Continental</u> <u>Nodes for KNIME</u> are a community extension that you can install by dragging the extension from the <u>KNIME Hub</u> to KNIME Analytics Platform or by installing the extension as described in this <u>video</u>.



Creating an XLS Control Table with Tag Values

As the saying goes, many roads lead to Rome. This section introduces two different roads or approaches for creating an XLS Control Table. (The second approach happens to be my personal favorite!) The "key node" in both examples is the <u>XLS Control Table Generator</u> node.

Options	Flow Variables Jo	b Manager Selection	Memory Policy
Operation Type (auto	matically set based o	n the provided input ta	ıble)
• from arb	itrary input table to XI	LS Control Table (wide	or long/unpivoted)
from long	g/unpivoted layout to	wide XLS Control Tab	e
Shift Rows Option			
	write colu	nn header to first row	
Result Table Structur	e Options		
unpiv	vot result table (for ea	sier post-processing a	nd re-pivoting)
	🗸 add additi	onal header columns	
		onal ficader coratinis	
	tion Strategy at Opera	ation Type 'long to wid	e'
Contradiction Resolu			(
	ontradicting informati	on? fail	`
	ontradicting informati	on? fail	
	ontradicting informati	on? fail	
	ontradicting informati		Cancel

Figure 3 The configuration dialog for the <u>XLS Control Table Generator</u> node

Approach 1: Table Creator + XLS Control Table Generator

The first approach to create an XLS Control Table with tags involves a combination of a <u>Table Creator</u> and an <u>XLS Control Table Generator</u> node. This is an easy approach, however the downside is that it entails a lot manual work creating the tag table and you have the problem that the tag table is static. Therefore, this approach is only recommended for small tables, where the number of rows and columns won't change.

Open the configuration window of the <u>Table Creator</u> node to add one or multiple tag values for each cell. If you want to enter multiple tags, remember to separate them with a comma. The <u>XLS Control Table</u> <u>Generator</u> node transforms the table into an XLS Control table and replaces the column names with letters and the row IDs with numbers. The checkbox "write column header to first row" gives you the option of retaining the column headers, similar to the option "add column headers" in the <u>Excel Writer</u> node.

Approach 2: XLS Control Table Generator + Rule Engine

The second approach we want to show uses a combination of two <u>XLS</u> <u>Control Table Generator</u> nodes with the unpivot option activated and a <u>Rule</u> <u>Engine</u> node.

Activate the checkbox "unpivot result table" in the configuration window of the first <u>XLS Control Table Generator</u> node to output a table that has one row for each cell in the input table, including the value, the row number, the column header, and more.

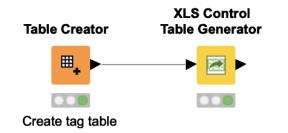


Figure 4 Here you can see one option to create a control table using the <u>Table Creator</u> node and the <u>XLS Control Table Generator</u> node

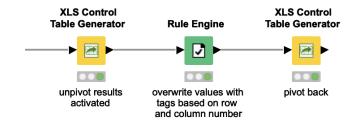


Figure 5 Here you can see another option to create a control table using a combination of two XLS Control Table Generator nodes and a Rule Engine node. The <u>Rule Engine</u> node is used to define tag values based on the row and column number.

								Rov	N ID	S Cell	S Column	S Colum	. Colum	S Colum	I Row	S RowID	S Value
								Row0		Al	A	00A	1	column1	1	Row0	Basic Report Example Mai 20
	Concate	nated table -	2:2226 - C	oncatenate				Row1		B1	В	00B	2	column2	1	Row0	?
ile Hilite	Navigation View							Row2		C1	С	00C	3	column3	1	Row0	?
					1			Row3		D1	D	00D	4	column4	1	Row0	?
	Table "default" - Rows: 8	Spec – Co	olumns: 7	Properties	Flow Var	iables		Row4		E1	E	00E	5	column5	1	Row0	?
								XLS Control		F1 G1	F C	00F 00G	5	column6 column7	1	Row0 Row0	2
Row ID	S column1	1. housed	S columns	S column4	S column	s s column	6 S column7	Table Generator		A2	A	00G	1	column1	2	Empty 0	?
Row0	Basic Report Example Mai 2019	?	?	?	?	?	?	Row8		B2	В	008	2	column2	2	Empty 0	7
Empty 0	?	?	?	?	?	?	?	Row9		C2	с	00C	3	column3	2	Empty 0	?
Row0_dup	Calendar Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Row1		D2	D	00D	4	column4	2	Empty 0	?
Row1	CW 18	?	?	2158.99	2486.56	2756.36	5486.35	Row1		E2	E	00E	5	column5	2	Empty 0	?
Row2	CW 19	2196.47	2963.47	2348.57	2786.57	2873.56	3947.56	Row1		F2	F	00F	6	column6	2	Empty 0	?
Row3	CW 20	2964.87	2145.78	3120.86	3452.99	3245.58	4279.26	Row1 Row1		G2 A3	6	00G 00A	1	column7 column1	2	Empty 0	Calendar Week
Row4	CW 21	2115.78	2657.23	2765.36	2275.96	1458.65	4823.45	unpivot results		B3	B	008	2	column2	3	Row0_dup	
Row5	CW 22	2678.36		2192.78	2571.45	3156.47	2			C3	c	00C	3	column3	3	Row0 dup	
Rows	CW 22	2078.30	1208.40	2192.78	2371.43	5150.47		activated Row1		D3	D	00D	4	column4	3	Row0_dup	Wednesday
								Row1		E3	E	00E	5	column5	3	Row0_dup	
								Row1		F3	F	00F	6	column6	3	Row0_dup	
								Row2		G3	G	00G	7	column7	3		Saturday
								Row2		A4	A	00A	1	column1	4	Row1	CW 18
								Row2 Row2		B4 C4	в	00B 00C	2	column2 column3	4	Row1 Row1	2
								Row2		D4	D	00C	4	column3	4	Row1	2158.99
								Row2		E4	E	00E	5	column5	4	Row1	2486.56
								Row2		F4	F	00F	6	column6	4	Row1	2756.36

Figure 6 Here you can see the resulting table when activating the checkbox "unpivot result table" in the <u>XLS Control Table Generator</u> node. The node creates one row for each cell including value, row number, column header, etc.

This table is a great basis to now transform values into tags with the <u>Rule</u> <u>Engine</u> node. For example, we can replace all values in the first row with the tag "header", or replace all values in the first column that have a row number higher than 3 three with the tag value "cw".

Hint: Activate the checkbox "Replace Column" and select the column "Value".

	Rule Editor Flow Variables Job Manager Selection	Memory Policy				Classified	values - 2:226	0 - Rule En	gine			
	Category Description		File Hilite	Navigation	View							
Column List ROWID	Category Description			Table "default" - Rows: 56 Spec - Columns: 8 Properties Flow Variables								
OWINDEX OWCOUNT			Row ID	S Cell	S Column	S Colum	Colum	S Colum	Row	S RowID	S Value	
Cell	Function		Row0	A1	A	00A	1	column1	1	Row0	header	
olumn	? < ?		Row1	B1	B	00B	2	column2	1	Row0	header	
olumn (comparable)	? <= ? ? = ?		Row2	C1	c	00C	3	column3	1	Row0	header	
olumn (number) olumn name	7 > 7		Row3	D1	D	00D	4	column4	1	Row0	header	
Row	? >= ?				E		5		1	Row0	header	
RowID	? AND ?				F				1	Row0	header	
Value					G		7		1	Row0	header	
					-		1		2	Empty 0	empty	
							2		2	Empty 0	empty	
	? XOR ?				c					Empty 0	empty	
	FALSE				D		4	column4	2	Empty 0	empty	
0			Row11	E2	E	00E	5	column5	2	Empty 0	empty	
w Variable List	NOT ?		Row12	F2	F	00F	6	column6	2	Empty 0	empty	
nime.workspace	Expression				G		7		2	Empty 0	empty	
	? 2 // \$double column name\$ > 5.0 => "large"	Row14	A3	Ā	00A	1	column1	3	Row0 dup	c_heade		
		l and blue"			В		2		3	Row0 dup	c_heade	
	2 > - ? ? <td?< td=""> <td?< td="" td<=""><td>3</td><td>Row0_dup</td><td>c_heade</td></td?<></td?<>	3	Row0_dup	c_heade								
			Row17	D3	D	00D	4	column4	3	Row0_dup	c_heade	
			Row18	E3	E	00E	5	column5	3		c heade	
			Row19	F3	F	00F	6	column6	3	Row0_dup	c heade	
				G3	G	00G	7	column7	3	Row0_dup	c heade	
			Row21	A4	А	00A	1	column1	4	Row1	cw, bord	
⊖ Ap	opend Column: prediction	S	Row22	B4	в	00B	2	column2	4	Row1	border	
A D	place Column: S Value		Row23	C4	с	00C	3	column3	4	Row1	border	
O Re			Row24	D4	D	00D	4	column4	4	Row1	border	
			Row25	E4	E	00E	5	column5	4	Row1	border	
			Row26	F4	F	00F	6	column6	4	Row1	border	
	OK	Apply Cancel										

Figure 7 On the left is the configuration dialog of the <u>Rule Engine</u> node. In the Expression section you can see defined rules to replace the original values with tags, based on the row and column number. On the right you can see the output table where the rules are applied, and the values are replaced with the different tags

A second <u>XLS Control Table Generator</u> node can transform this table back into its original form, where the values are replaced with the different tag. This feature is automatically activated when the node detects an input table that was created by an <u>XLS Control Table</u> <u>Generator</u> node in unpivot mode.

This approach involves much less manual work compared to the first approach and can be implemented in a way to handle changing table dimensions gracefully.

Hint: Another helpful node to create a static XLS Control Tables is the <u>XLS Control Table</u> from Cell Range node.

Adding Formatting Actions based on Tag Values

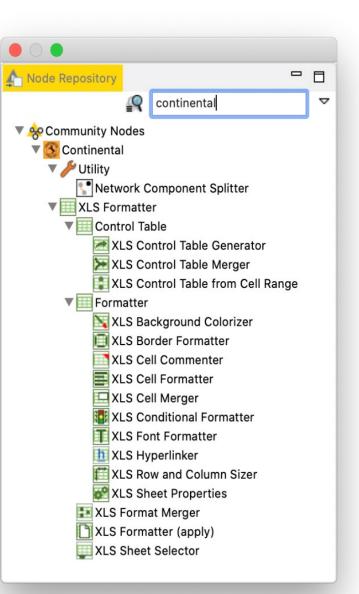
The next step to adding background colors or a border, etc. to your table is a sequence of <u>XLS Formatting</u> nodes, similar to the workflow in figure 2.

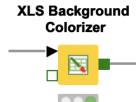
As you can see all nodes in the example workflow have two input ports and one output port:

- a data input port
- an optional XLS Formatter input port (square with green border)
- an output port, which is an XLS Formatter port

The green square is a special port type of the extension, which collects the different formatting instructions. The data input port expects the table with the tag values. The optional input port can be used to feed an XLS Formatting table with previous formatting instructions to which the instructions of the node should be added.

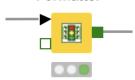
The figure on the right shows you an overview of all the nodes in the Continental extension. I will introduce my favorite ones and leave it up to you to explore the others.





The <u>XLS Background Colorizer</u> node changes the background color of cells. You can assign either a static color and / or pattern fill. One option is to assign the same color to all cells with a specific tag value, e.g. all cells with tag "header" should have a yellow background. Another option is to use RGB values in either hex syntax #FFD800 or decimal syntax R/G/B as tags and use them as the background color.

XLS Conditional Formatter



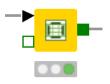
The <u>XLS Conditional Formatter</u> node changes the background for the cells with a certain tag value according to their numerical value. In the configuration window you can define a color scale by setting a minimum and maximum value and assigning a color to each. Optionally you could set a mid point value and assign a color to that. Cells with values higher or lower than the thresholds will have the background color of the minimum /maximum value.

XLS Sheet Selector



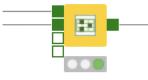
The <u>XLS Sheet Selector</u> and the <u>XLS Merger</u> node are really helpful nodes if your EXCEL file has more than one sheet. By default the formatting is always applied to the first sheet. So, if you have an Excel file with only one sheet you don't have to worry about these two nodes. However, if you have multiple sheets the <u>XLS</u> <u>Sheet Selector</u> allows you to define which sheet your XLS Control table is for.

XLS Border Formatter

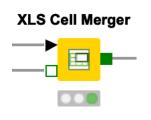


The <u>XLS Border Formatter</u> node can add borders to a given range specified by a certain tag or by all tags. By activating the corresponding checkboxes, you can add borders to the top, right, bottom, and left. In addition to adding a border around the range specified by the tags, the node gives you the option to use inner vertical and horizontal boarder lines in each cell, too.

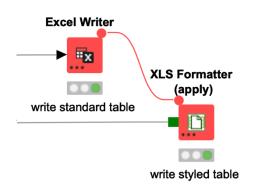
XLS Format Merger



The <u>XLS Format Merger</u> node allows you to either combine formatting instructions for different sheets prior to using the <u>XLS Formatter (apply)</u> node or when applied to the same sheet, it merges the properties at the lowest detail level (e.g. the formatting instructions for the cell A1 is bold in control table one and italic in control table two. The subsequent formatting instruction for A1 is italic and bold). Thereby, the upper input port overwrites a lower one in case of conflicting information (e.g. two different font colors for the same cell).



The <u>XLS Cell Merger</u> node merges the cells for given rectangular ranges of input tags into one cell. For example, we can merge all cells in the first row and centralize the title with the XLS Font Formatter node. This node works only on strictly rectangular ranges. The value of the merged cell is the value of the most top left cell of the merged range.



The <u>XLS Formatter (apply)</u> node reads an unformatted Excel file, applies all the collected formatting instructions, and saves the nice Excel file in the defined output location.

This was a short introduction. You can find further information about the different XLS Formatter nodes in the Continental extension in the documentation <u>https://www.knime.com/community/continental-nodes-for-knime-xls-formatter</u> or, from within KNIME Analytics Platform, by looking in the node description of each individual node.



The KNIME Booklet for Excel Users

Are you an experienced Excel user and want to start using KNIME Analytics Platform?

It's sometimes difficult to switch from one software tool to another. But this booklet is the perfect starting point as it maps the most commonly used Excel functions and techniques to their KNIME equivalents. Find out, for example, how data reading, filtering, sorting, and vlookup work in KNIME.

For a complete introduction to KNIME, please refer to my book "KNIME Beginner's Luck" available from KNIME Press under <u>https://www.knime.com/knimepress</u>

About the Author

Kathrin Melcher is currently a Data Scientist at KNIME. She holds an MSc in Mathematics, from the University of Konstanz, Germany. She joined the KNIME Evangelism team in May 2017 and has a strong interest in data science, machine learning, and algorithms. She enjoys teaching and sharing her knowledge on these topics.