

[L2-DW] KNIME Analytics Platform for Data Wranglers: Advanced

April 17, 2023

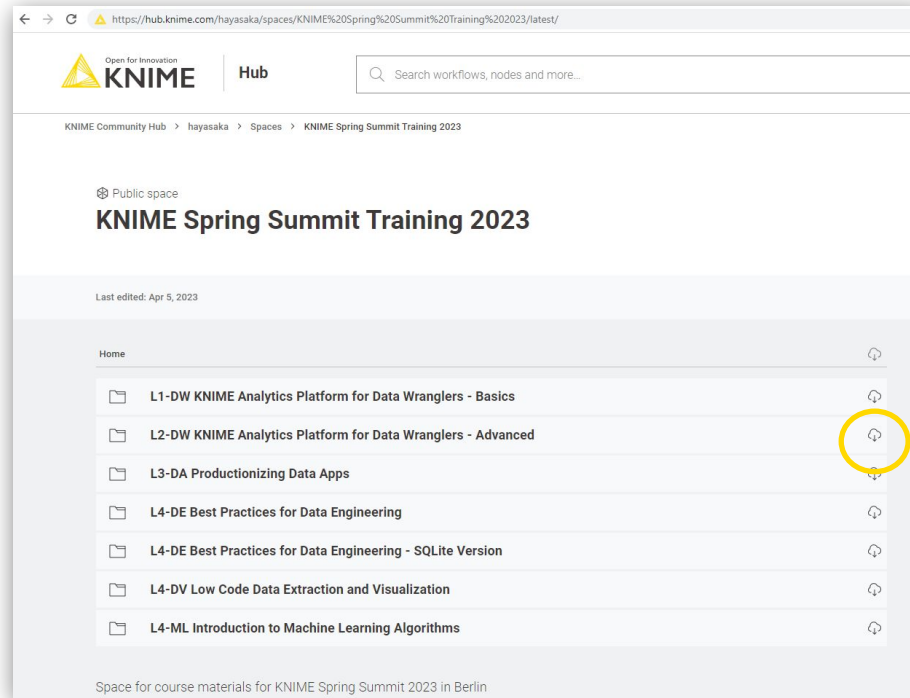


Structure of the course

- This course consists of five sessions
 - Data/Time and Data Export
 - Flow Variables and Shared Components
 - Workflow Control: Loops, Switch, and Try-Catch
 - Formatting Excel Tables and Introduction to Data Science
 - Q&A and Summary
- Structure of each session
 - Introduction to the topic
 - Hands-on exercise
 - Solution walk-through

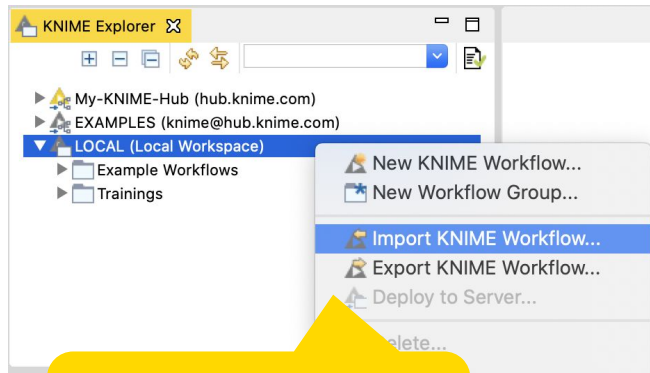
Downloading the Exercises

- Download the course material from the KNIME Community Hub
<https://hub.knime.com/hayasaka/spaces/KNIME%20Spring%20Summit%20Training%202023/latest/>

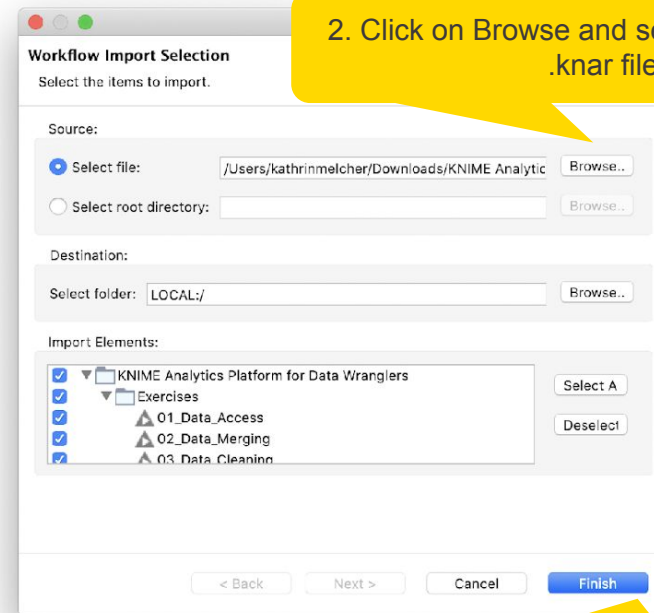


Importing the Exercises

- Import the course material to KNIME Analytics Platform



1. Right click on LOCAL and select Import KNIME Workflow....



2. Click on Browse and select downloaded .knar file

3. Click on Finish

Date/Time Data



Date & Time Overview

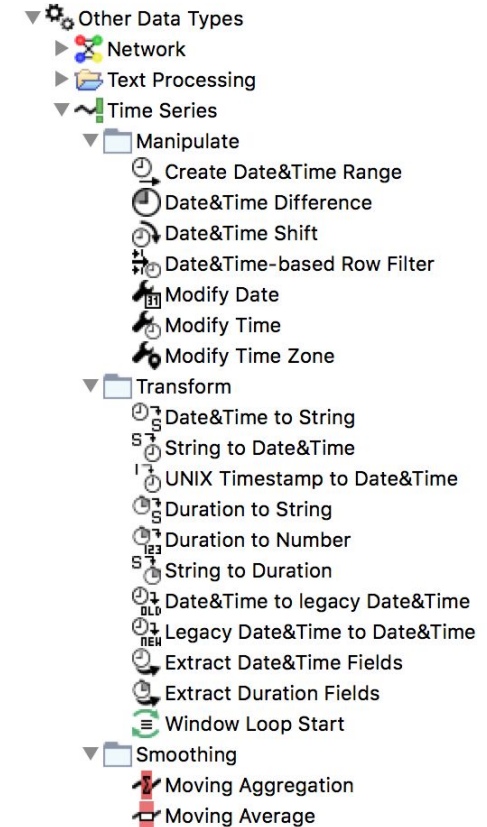
- Dedicated data type for date and time data
- Supported in Date&Time nodes
 - (and others: GroupBy, Pivot, Line Plot)
- Complete re-write in KNIME 3.4

Output table - 2:46 - String to Date&Time

File Hilite Navigation View

Table "default" - Rows: 524160

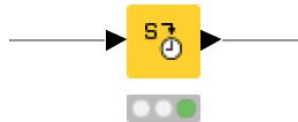
Row ID	timestamp	Intensity
Row0	2007-01-01T00:01	10.4
Row1	2007-01-01T00:02	10.4
Row2	2007-01-01T00:03	10.4
Row3	2007-01-01T00:04	10.4
Row4	2007-01-01T00:05	10.4
Row5	2007-01-01T00:06	10.4
Row6	2007-01-01T00:07	10.2
Row7	2007-01-01T00:08	10.2
Row8	2007-01-01T00:09	10.2
Row9	2007-01-01T00:10	10.2
Row10	2007-01-01T00:11	10.2
Row11	2007-01-01T00:12	10.2
Row12	2007-01-01T00:13	10.2
Row13	2007-01-01T00:14	10.2
Row14	2007-01-01T00:15	10.2



String to Date&Time

- Converts date/time data from string into a native Date&Time cell
- Guesses correct format for many types of date formatting
 - Enter format manually if auto-guessing didn't work
 - KNIME automatically adds custom formats to auto-guess list
 - Converts multiple columns of same date format in one node

String to Date&Time



Select type of output column

The screenshot shows the 'String to Date&Time' dialog box with several annotations:

- Select columns to transform:** Points to the 'Include' list containing 'Date'.
- Enter date format manually:** Points to the 'Date format' dropdown set to 'M-d-yyyy'.
- Click to auto-guess format:** Points to the 'Guess data type and format' button.

The dialog box includes tabs for 'Options', 'Flow Variables', 'Job Manager Selection', and 'Me'. It has radio buttons for 'Manual Selection' (selected) and 'Wildcard/Regex Selection'. The 'Exclude' list contains various columns like City, Country, CustomerID, etc. The 'Replace/Append Selection' section has 'Replace selected columns' selected. The 'Type and Format Selection' section shows 'New type: Date', 'Locale: en-GB', and 'Content of the first cell: 11-2-2014'. At the bottom, there are 'OK - Execute', 'Apply', and 'Cancel' buttons, along with a 'Fail on error' checkbox.

Date&Time – Data Types

Output table - 2:50 - Create Date&Time Range

File Hilite Navigation View

Table "default" - Rows: 1000

Row ID	Date
Row0	2017-01-19
Row1	2017-01-19
Row2	2017-01-20
Row3	2017-01-20
Row4	2017-01-20
Row5	2017-01-21
Row6	2017-01-21
Row7	2017-01-22
Row8	2017-01-22
Row9	2017-01-22
Row10	2017-01-23

Date

Output table - 2:50 - Create Date&Time Range

File Hilite Navigation View

Table "default" - Rows: 1000

Row ID	Date&Time
Row0	2017-01-19T13:00:46
Row1	2017-01-19T21:46:57
Row2	2017-01-20T06:33:08
Row3	2017-01-20T15:19:20
Row4	2017-01-21T00:05:31
Row5	2017-01-21T08:51:42
Row6	2017-01-21T17:37:53
Row7	2017-01-22T02:24:04
Row8	2017-01-22T11:10:15
Row9	2017-01-22T19:56:27
Row10	2017-01-23T04:42:38

Date & Time

Dialog - 2:50 - Create Date&Time Range

Options Flow Variables Job Manager Selection Memory Policy

Output Settings

Output type:
 ☒ Date&time
 ☐ Date&time with zone

New column name: Date&Time

Mode Selection

Number of rows: ☒ Fixed: 1,000
 ☐ Variable

Starting Point

Start: Date: 2017-01-19 Time: 13:00:46
 Time Zone: Europe/Berlin

☐ Use execution date&time

Ending Point

☐ Interval:
 ☒ End: Date: 2018-01-19 Time: 14:00:46
 ☐ Use execution date&time

OK Apply Cancel ?

Output table - 2:50 - Create Date&Time Range

File Hilite Navigation View

Table "default" - Rows: 1000

Row ID	Time
Row0	14:02:31.155
Row1	14:02:31.155
Row2	14:02:31.155
Row3	14:02:31.155
Row4	14:02:31.155
Row5	14:02:31.155
Row6	14:02:31.155
Row7	14:02:31.155
Row8	14:02:31.155
Row9	14:02:31.155
Row10	14:02:31.155

Time

Output table - 2:50 - Create Date&Time Range

File Hilite Navigation View

Table "default" - Rows: 1000 Spec - Column: 1

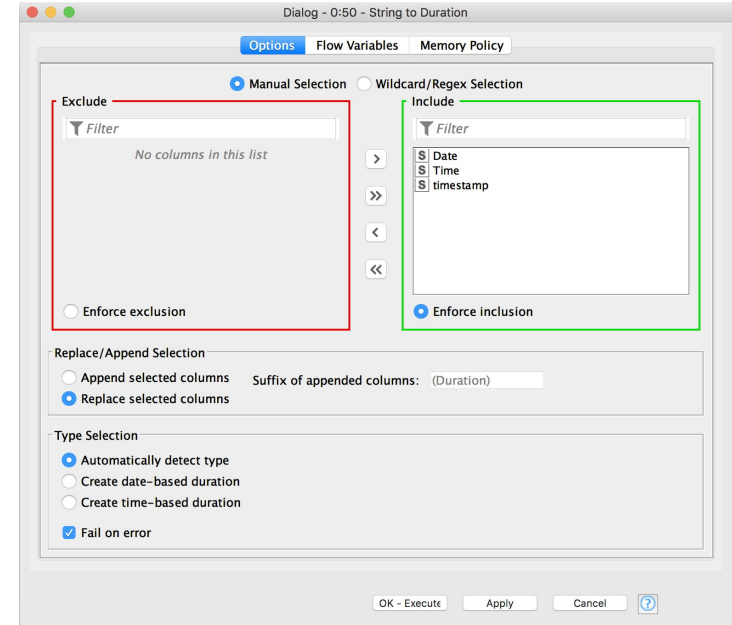
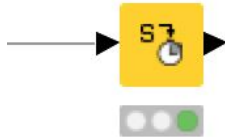
Row ID	Time
Row0	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row1	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row2	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row3	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row4	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row5	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row6	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row7	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row8	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row9	2018-01-19T14:02:31.155+01:00[Europe/Berlin]
Row10	2018-01-19T14:02:31.155+01:00[Europe/Berlin]

Date & Time +
Time zone

String to Duration

- Takes a string and converts it to a duration cell
 - Three different options to format input strings
- Example: Convert 1 year, 2 months, 3 weeks, and 4 days to duration cell
 - ISO-8601: “P1Y2M3W4D”
 - Short letter: “1y 2M 3w 4d”
 - Long word: “1 year 2 months 3 weeks 4 days”

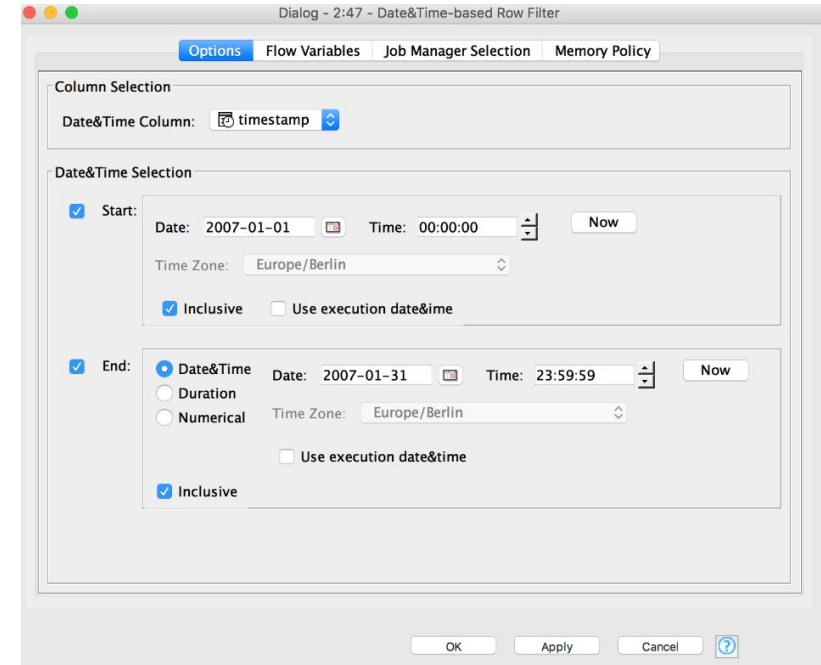
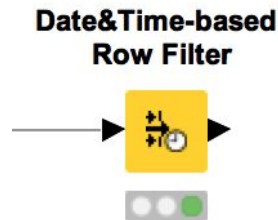
String to Duration



Row ID	iso	short	long	iso(Duration)	short(Duration)	long(Duration)
Row0	P1Y2M3W4D	1y 2M 3w 4d	1 year 2 months 3 weeks 4 days	1y 2M 25d	1y 2M 25d	1y 2M 25d

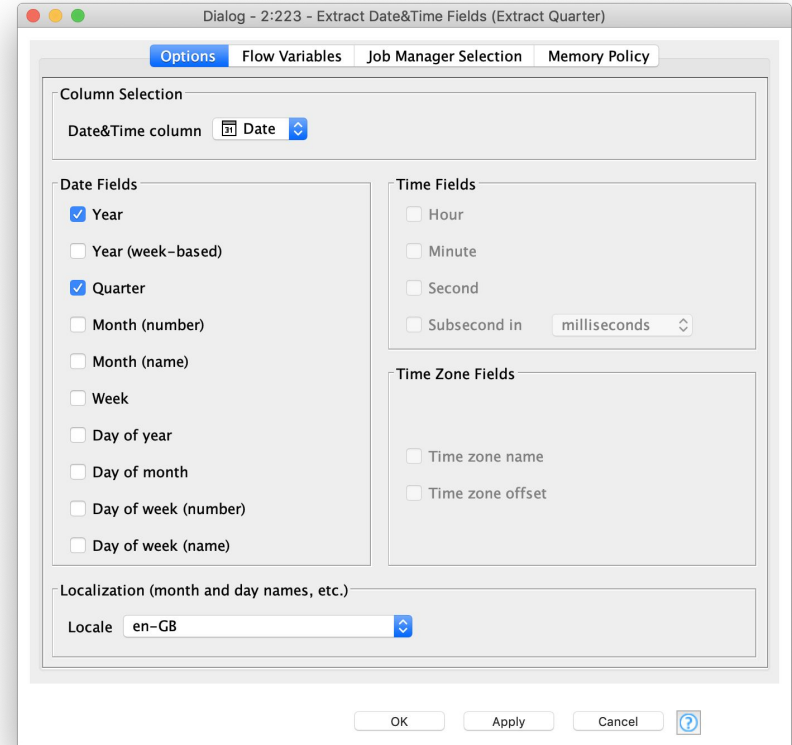
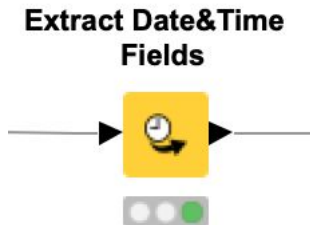
Date&Time-based Row Filter

- Filters rows from a specified time period
- Range can be limited on upper bound, lower bound or both
- Options for end point:
 - Date&Time: Fixed data and time
 - Duration: Duration string (e.g. 2y 3M)
 - Numerical: Select granularity from dropdown and enter number



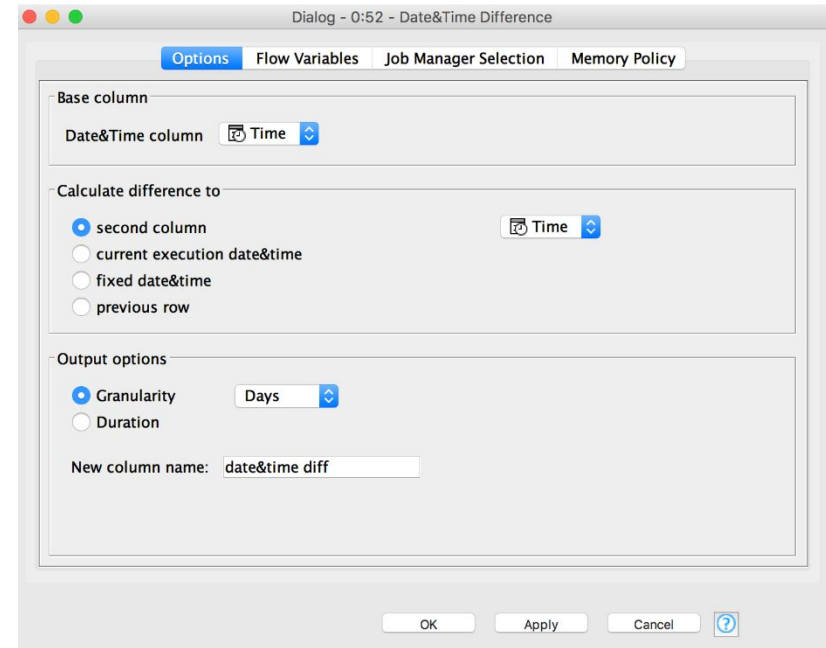
Extract Date&Time Fields

- Extracts date fields (year, day, month) or time fields (hour, minute, second) from a date&time cell.
- You can pick and choose which fields to include
- Useful when used in combination with data aggregation nodes (groupby, pivot etc.)



Date&Time Difference

- Check the difference between a time column and...
 - Another time column
 - Execution time
 - User-defined time
 - Time from previous row
- Choose desired resolution (days, hours, minutes, etc.)

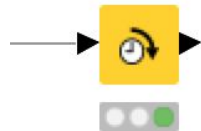


To calculate difference to second column, both columns need to have the same type!

Date&Time Shift

- Shifts date or time by either a duration or a numerical value
- Use duration:
 - Use duration column
 - Or shift by user-defined value
 - E.g. 1y, 2M, 5h, etc.
- Use numerical in combination with user-defined granularity
 - Use numerical column
 - Or shift by user-defined value

Date&Time Shift

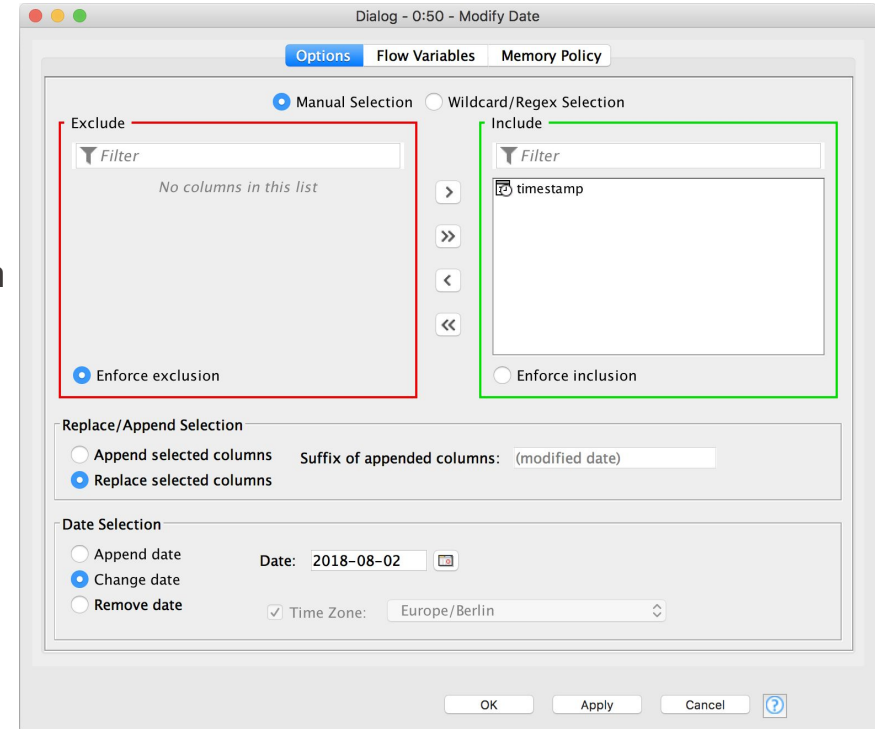
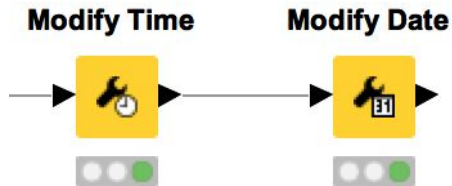


Select granularity

The screenshot shows the 'Date&Time Shift' dialog box with the 'Manual Selection' tab active. The 'Exclude' section on the left is empty, while the 'Include' section on the right contains the 'timestamp' column. Below these are sections for 'Replace/Append Selection' (with 'Replace selected columns' selected), 'Shift Value Selection' (with 'Use Duration' selected), and 'Use Numerical' (with 'Numerical value' selected and a value of '1' entered). The 'Granularity' is set to 'Years'. The dialog has tabs for 'Options', 'Flow Variables', 'Job Manager Selection', and 'Memory Policy' at the top, and 'OK', 'Apply', 'Cancel', and a help icon at the bottom.

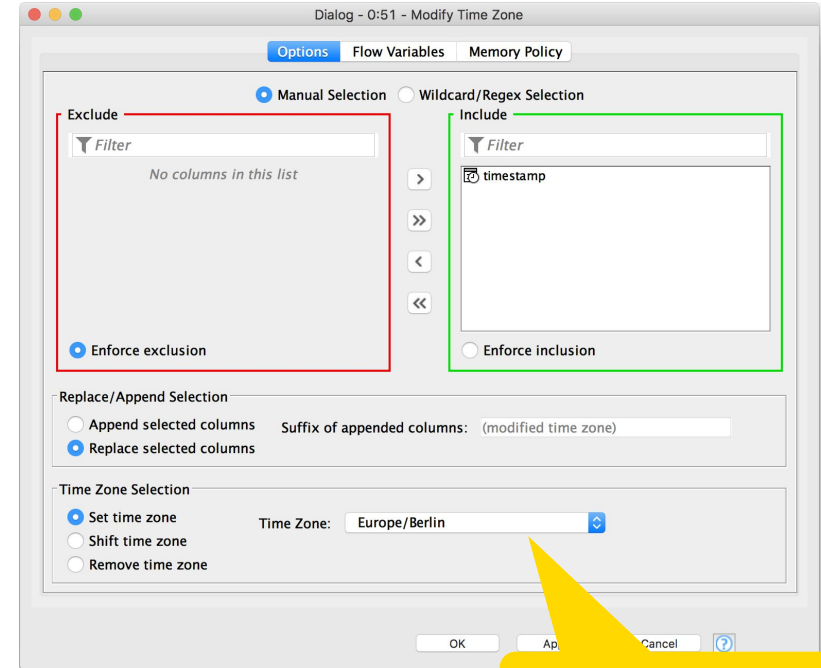
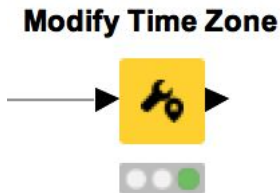
Modify Time / Modify Date

- Modify Date&Time columns
- Three options:
 - Appends time (date) to date (time) column
 - Changes time (date) to fixed value
 - Removes time (date) from Date&Time column
- Column selection shows only columns suitable for currently selected option



Modify Time Zone

- Similar to Modify Time/Modify Date
- Input: Date&Time
 - Set time zone
- Input: Date&Time (Time zone)
 - Set time zone
 - Shift time zone
 - Remove time zone



Select time zone from dropdown list

Exercise: 08_DateTime_Manipulation

- Convert order dates from string to Date&Time
- Extract the product purchases that were submitted in 2019
- Extract the remaining product purchases into a separate table
- Extract monthly and year of each product purchase into separate columns
- Plot monthly sales in a line plot

Exporting Data & Deployment



Exporting Data

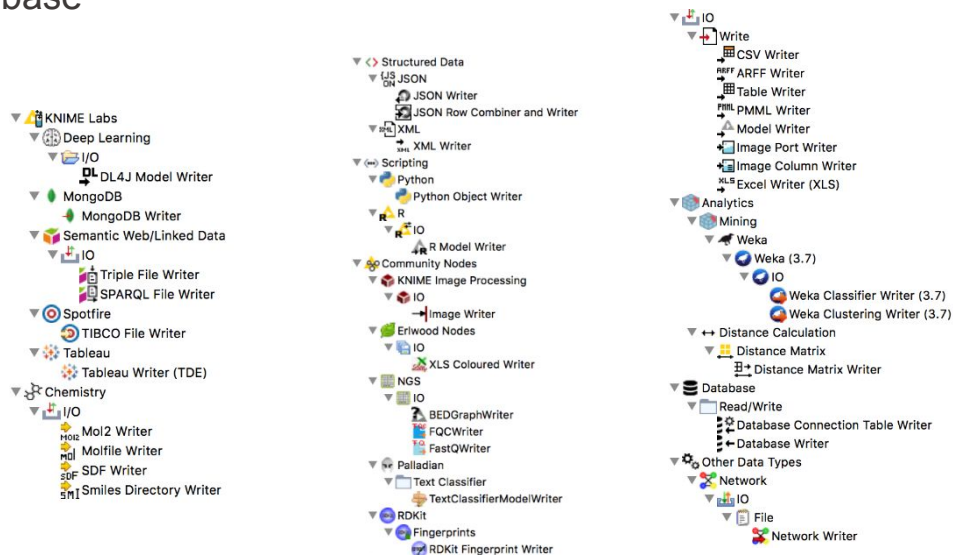
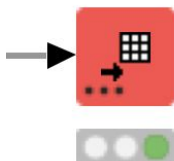
After an analysis is completed, what next?

- Write results to a file
- Upload results to a Cloud Storage
- Create/update a database
- Generate a rich report using BIRT
- Send your data to Tableau, Spotfire, PowerBI to create a report
- Deploy via KNIME KNIME Business Hub
- Deploy your model as RESTful web service

Data Export Nodes

- Typically characterized by:
 - Magenta color
 - 1 input port, no output ports
 - Create file on file system or write to database

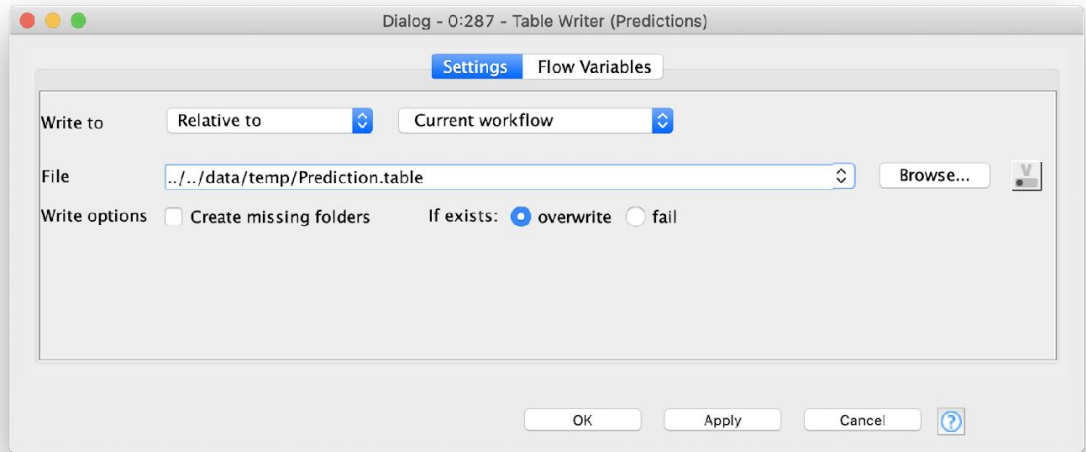
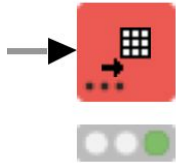
Table Writer



<https://www.youtube.com/watch?v=Og7VZOJhsOc&feature=youtu.be>

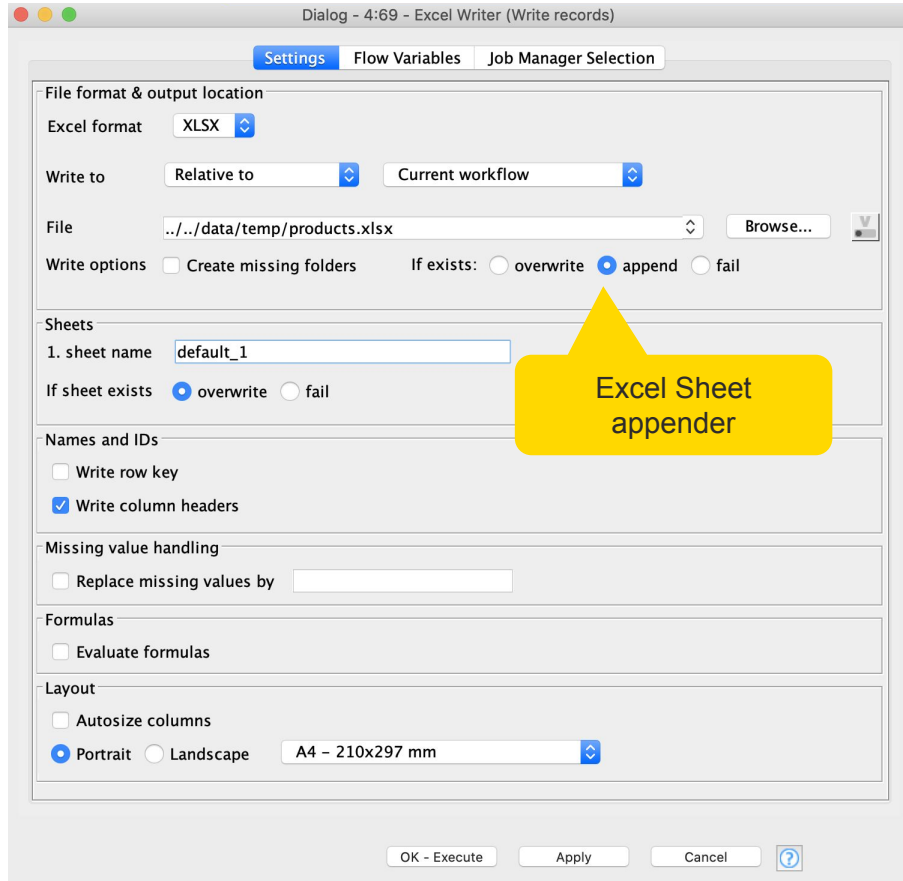
Table Writer

Table Writer



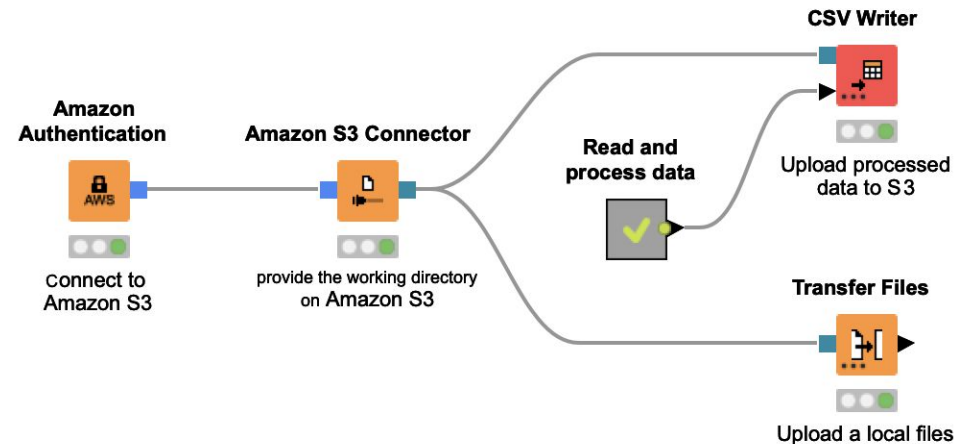
Excel Writer

- Writes the input table into a spreadsheet of an Excel file
- Select append, to append a spreadsheet to an existing Excel File and define the name of the new sheet

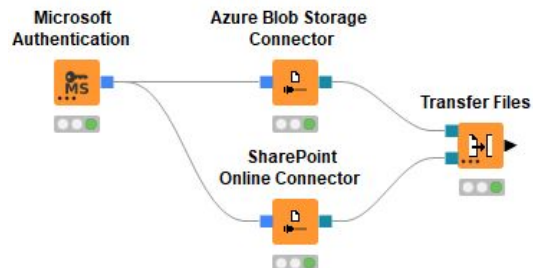


Write Files to a Remote File System

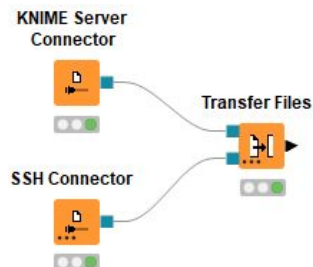
- The new file handling framework makes it easy to upload data to remote file systems
 - Write processed data directly with a writer node
 - Upload local files with the Transfer Files node
- Supported file systems
 - Microsoft Azure
 - Google
 - Amazon
 - Databricks
 - BigData file systems (hdfs, httpFS, ...)
 - On-premise (e.g. ssh, ftp, ...)



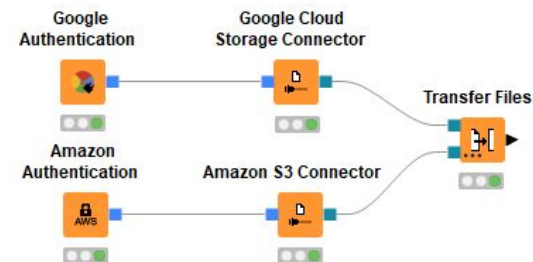
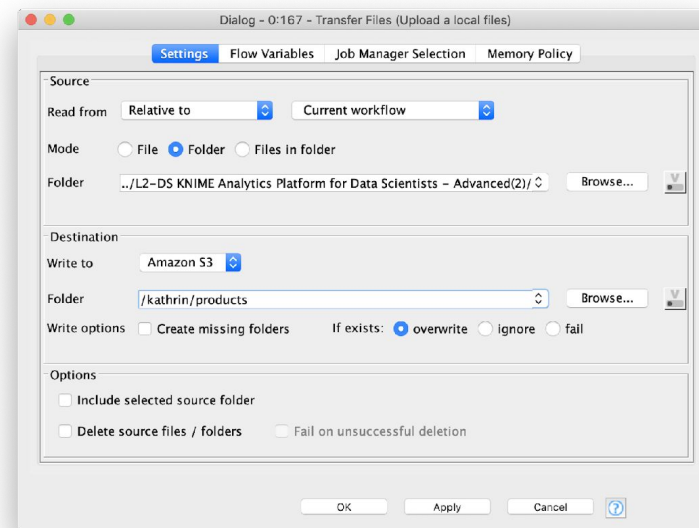
Full Flexibility with the Transfer Files node



Same cloud environment



On-premise

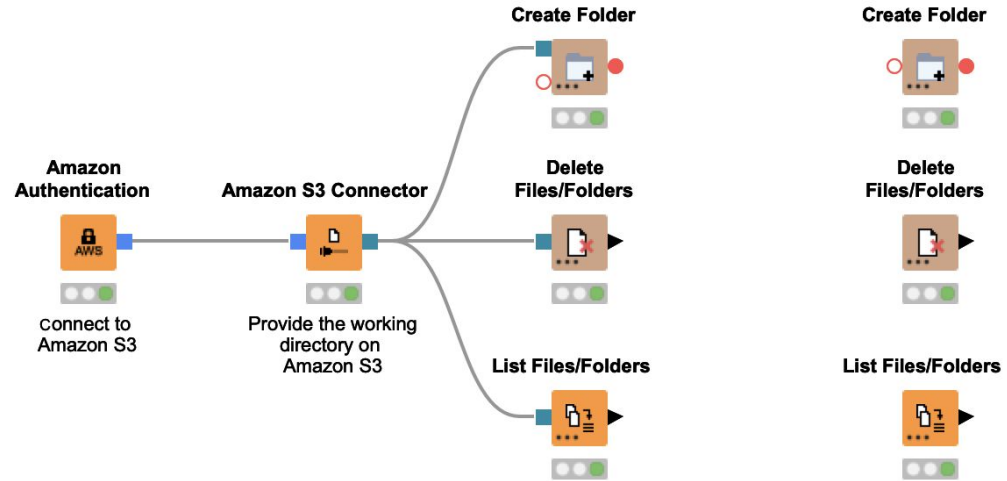


Cross cloud environments

Other Utility Nodes

Can be used with local and remote file systems

- Create a folder
- Delete files or folders
- List all files in a folder

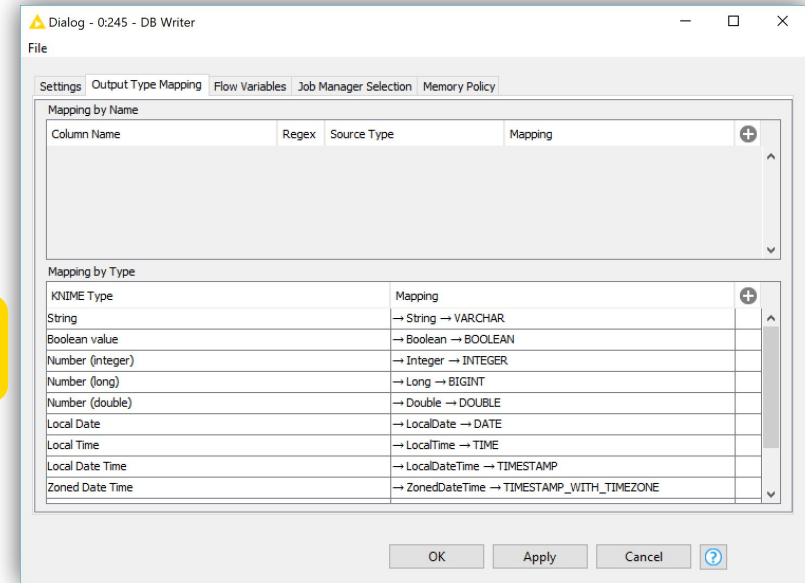
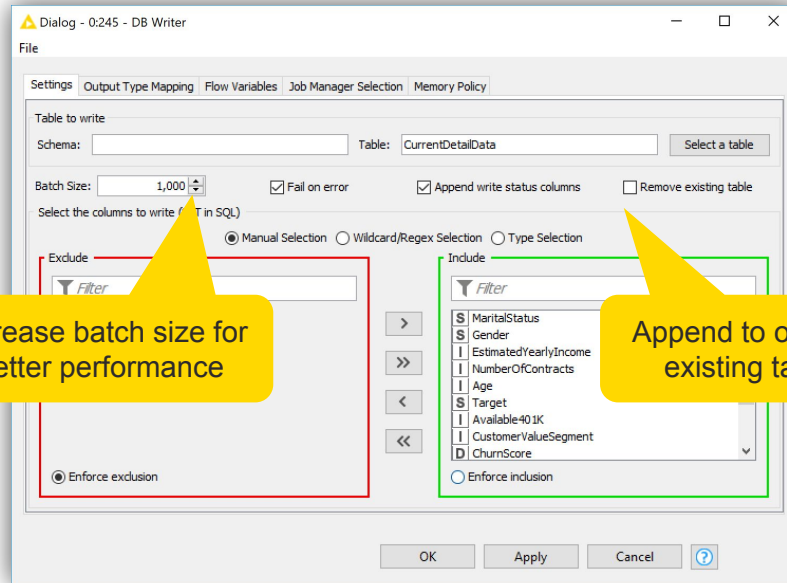
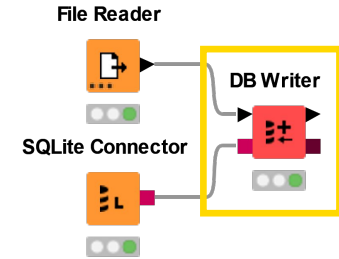


- Further information about file handling

https://docs.knime.com/latest/analytics_platform_file_handling_guide/index.html

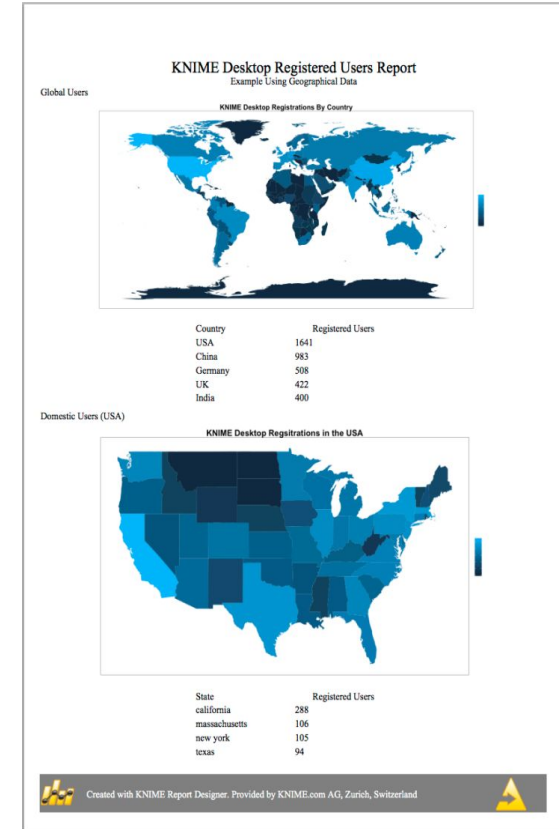
DB Writer

- Writes data from a KNIME data table **directly** into a database table



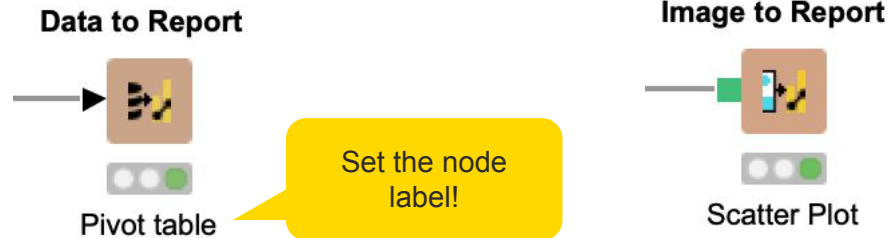
Reporting in KNIME

- Reporting in KNIME is done via a 3rd party application named BIRT (Business Intelligence Reporting Tool)
- Data is sent to BIRT from KNIME using special nodes.
- Reports in BIRT are constructed from report items, which may include images, tables, charts and labels.
- Reports may be generated in a variety of formats (html, pdf, pptx, xlsx, docx, ...)



Send Image / Data to Report

- Sends a data table or images to BIRT
- PNG and SVG are supported formats (see node description for details)
- Hint: The node label will be used to identify the data source in the reporting view
-> Make sure to use fitting labels if you have more than one data source



Edit the Report

Open the workflow > Click the Report Editor button in the tool bar

The screenshot displays the KNIME Analytics Platform interface. A yellow arrow points to the 'Report Editor' button in the top toolbar. The main workspace shows a workflow titled 'Activity I: Exporting Data' with nodes: Fully Joined Data, Partitioning, Decision Tree Learner, Decision Tree Predictor, Scorer, Normalizer, Heatmap, Image to Report, Row Filter, Column Filter, and Data to Report. A yellow box highlights the 'Data to Report' node. The right sidebar shows the 'Data to Report' dialog with the following content:

Activity I: Exporting Data

- Write predictions to disk as a KNIME table.
- Write Model to a PMML model file
- Create a heatmap of the normalized confusion matrix for your model and send it to BIRT.
- Send your model accuracy to BIRT.
- Define a very simple report showing the model accuracy and the heatmap of the confusion matrix.
- Generate a PDF of your report.

Data to Report

The incoming data is provided as a data set to the KNIME Report Designer.

Dialog Options

Use custom image scaling

If checked, images in the table will be scaled according to the specified width and height settings. Otherwise the default size of the involved renderer is used.

It is recommended to try this option if the image quality in the report does not meet the expectations. Depending on the renderer scaling the images here may provide better results.

Export Image As

Sends any image or rendered graphics contained in the table in the selected format to BIRT (the reporting library). PNG is the default and will result in a pixel graphic without further modification in the report.

SVG is currently in a development stage and is supported in BIRT only with few modifications to the report. These are the steps:

- Select SVG in this dialog, possibly change the resolution in the image scaling fields.
- Switch to the reporting, make sure nothing is selected in the report editor and switch to the "Script" tab. Select the "beforeRender" script and enter this command "KNIME.enableSVGImagesInPDF(reportContext)" (no quotes).
- Switch back to main report editor (tab "Layout"), design the report and add the dynamic image by linking it to the KNIME data source.
- Select the dynamic image report element and choose in the "Property Editor - Image" view under "Properties" the tab "Advanced" (note, you might need to scroll). In the property list search for "Type expression" and edit the entry. Enter "image/svg+xml", including the quote characters.

Height

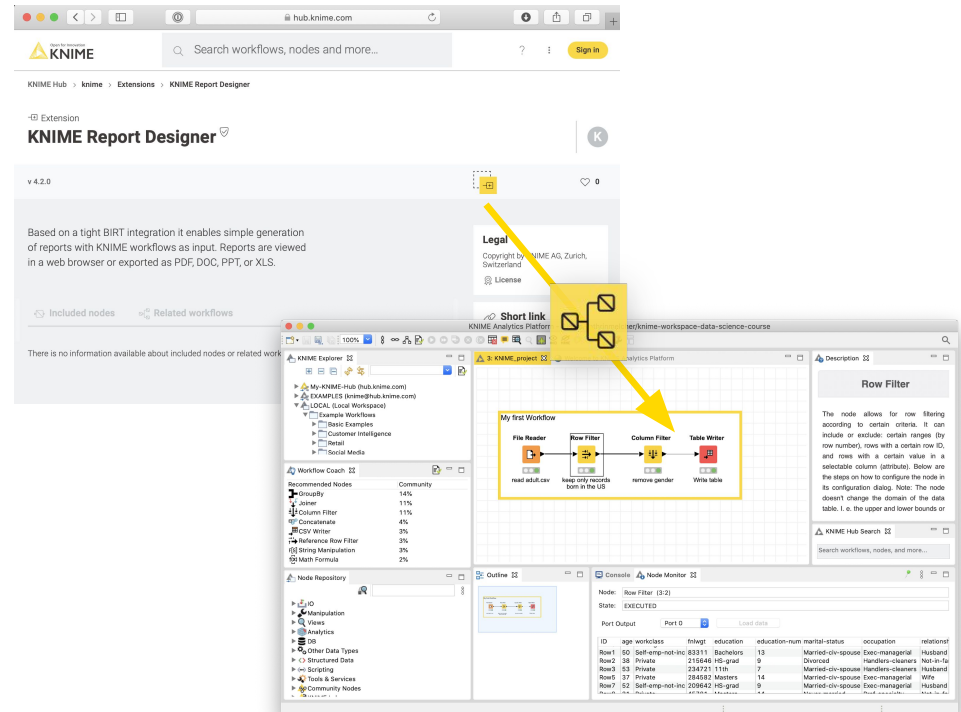
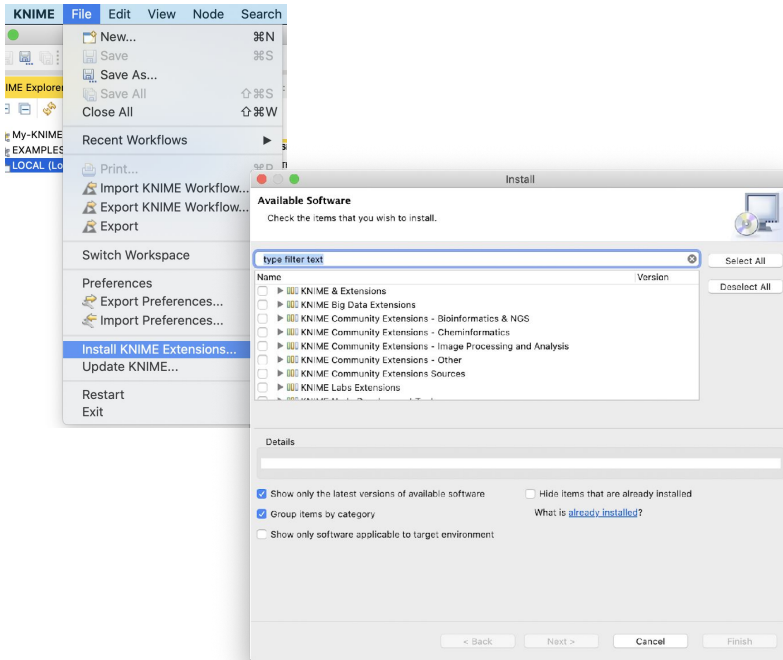
The custom height of the image.

KNIME Console

```
*****
*** Welcome to KNIME Analytics Platform v4.0.0.v201906260931 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
```


Installing Extensions

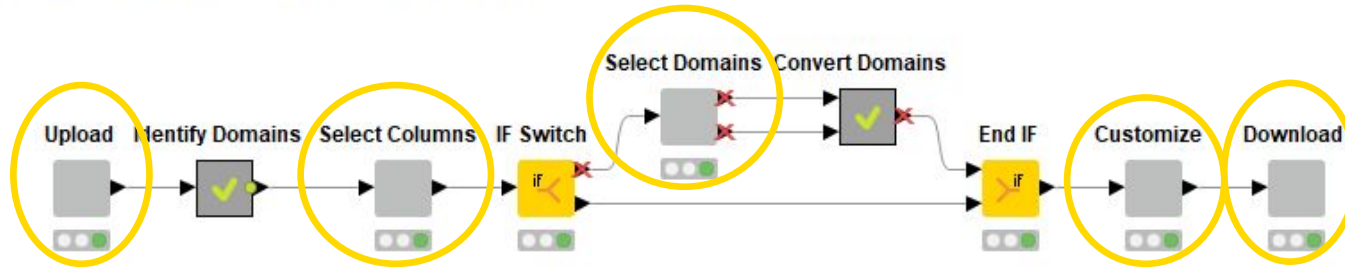
- Install the KNIME Report Designer Extension to use BIRT
- Install extension by going to File -> Install KNIME Extension or via Drag & Drop from the KNIME Community Hub



Creating a Data App on KNIME Business Hub

The Process Step by Step

1. Upload your data / Select one of the available datasets
2. Select the columns to visualize (maximum 3)
3. Convert the domain of the columns (OPTIONAL)
4. Customize the visualizations interactively
5. Download the images of the customized charts



Step 1
Upload File

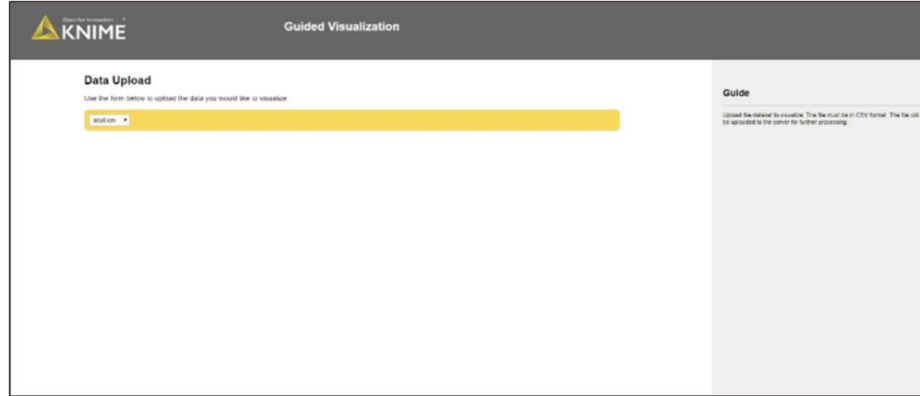
Step 2
Select Columns

Step 3
Customize Column
Domains

Step 4
Interactive View

Step 5
Download Image

Execute Workflow as Data App on KNIME Business Hub



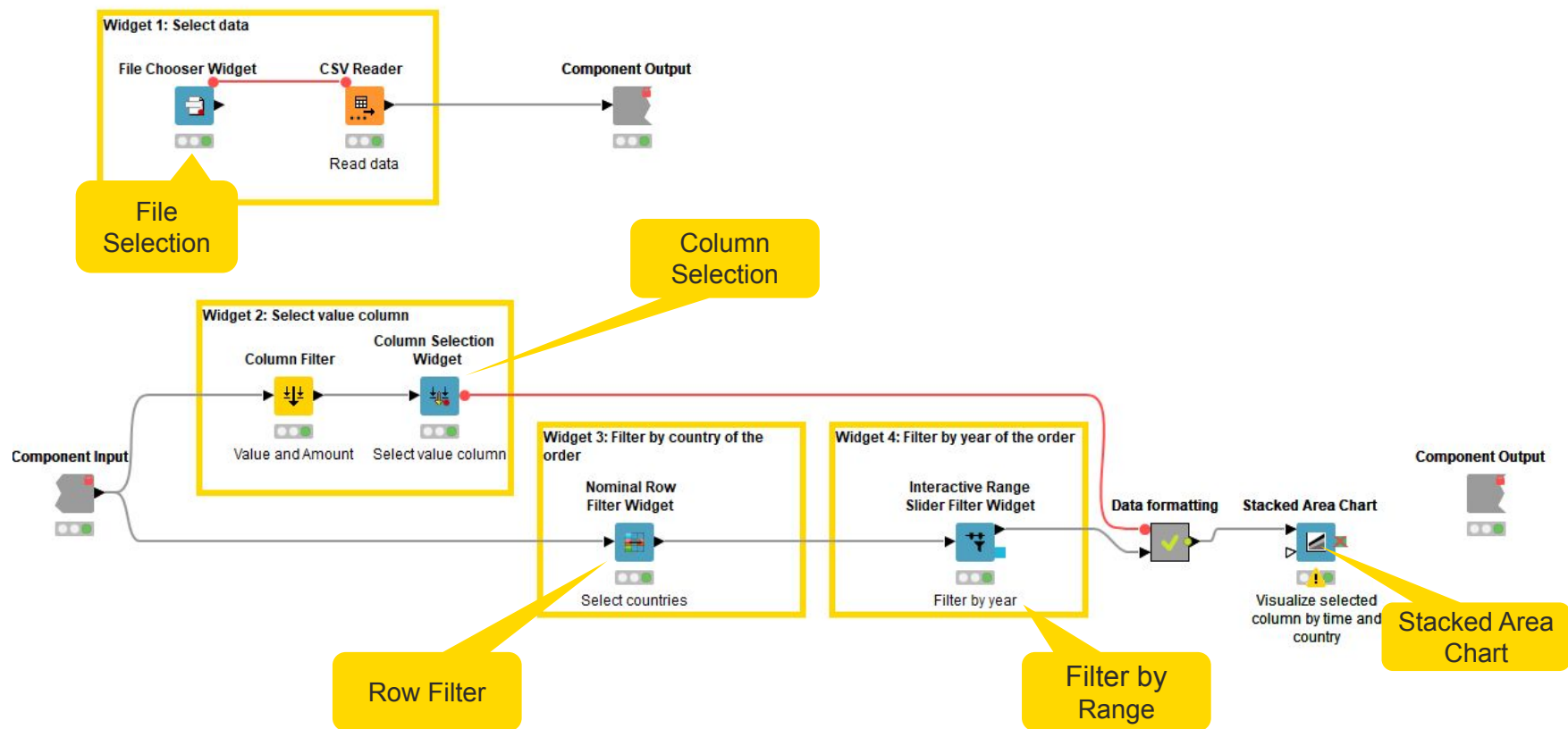
**Data App Page
(Step 1)
Upload File**

**Available in
KNIME
Business Hub**



**Data App Page
(Step 4)
Interactive View**

Components to Produce Dashboard on Web Page



Exercise: 09_Deployment

- Write the clean customer data to an Excel file into the folder "data/temp"
- Write the full transaction data to the "Transactions.sqlite" database

Flow Variables

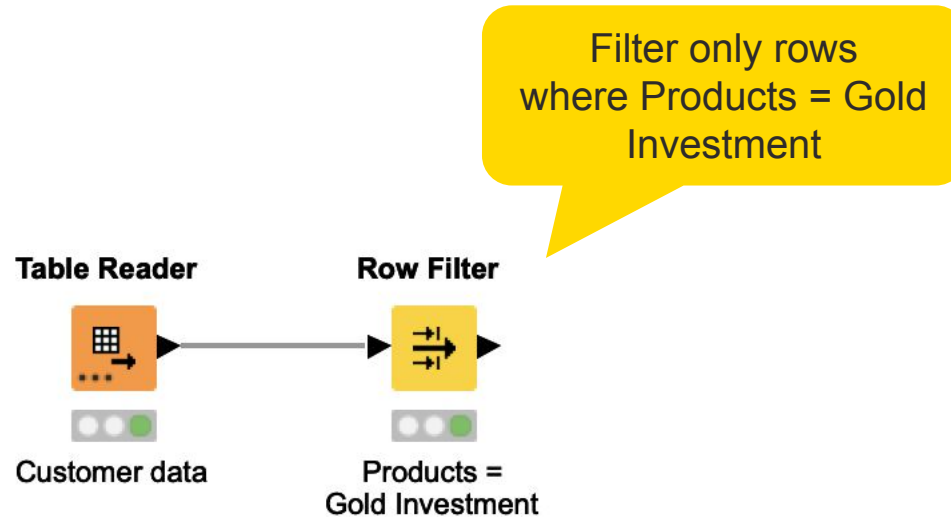


Goal of this Session

- What is a Flow Variable?
- Create a Flow Variable
- Use a Flow Variable as a parameter in the node settings
- Use a Configuration node to parameterize a Component
- Use a Widget node to enable interaction on a Data App page

Flow Variables: Usage Example

- Each month you need to produce a sales report for the most popular product



Flow Variables: Usage Example

- Each month I need to launch the Analytics Platform, aggregate the data to identify the most popular product, and update the Row Filter accordingly
- Or do I? Maybe Flow Variables can help...

Automatically Filter by Most Popular Product

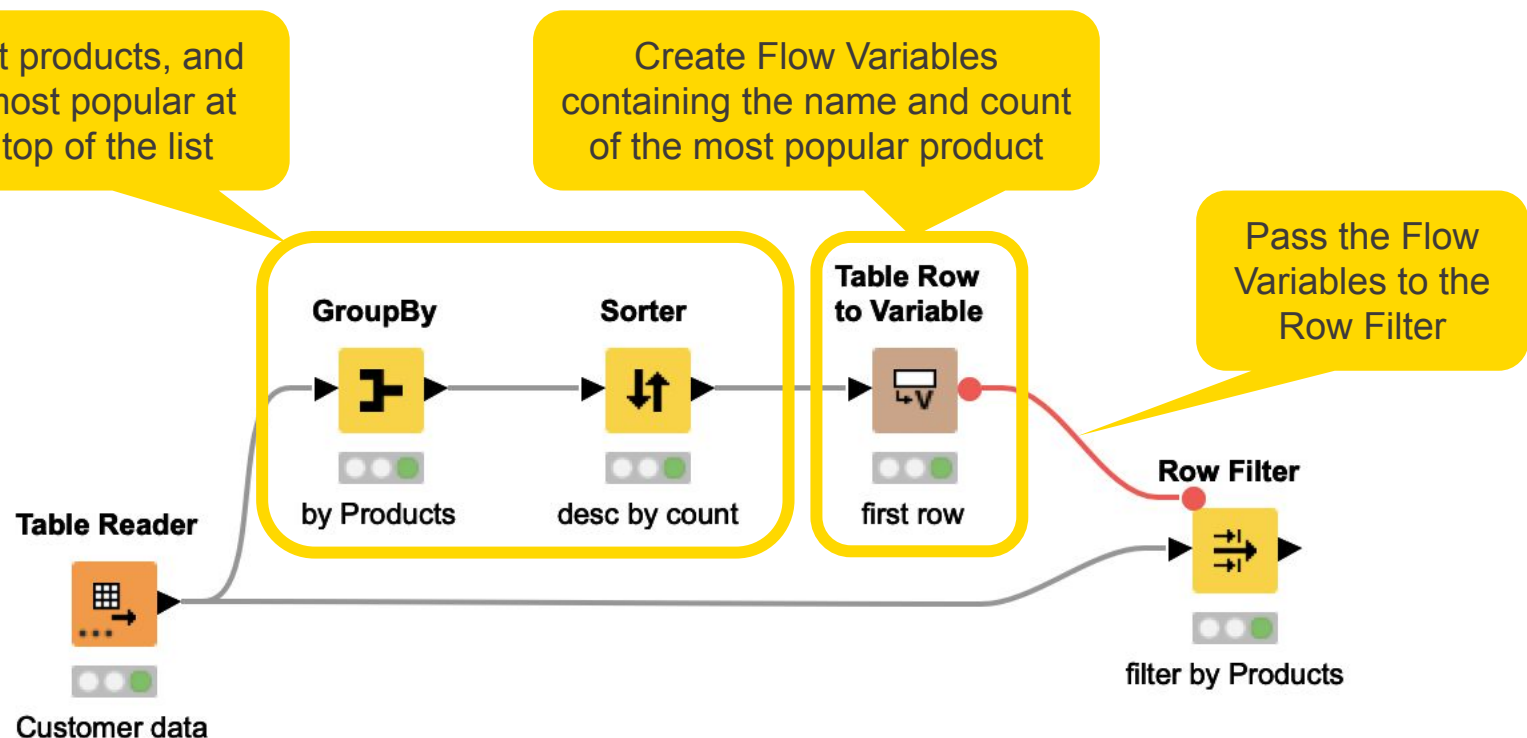
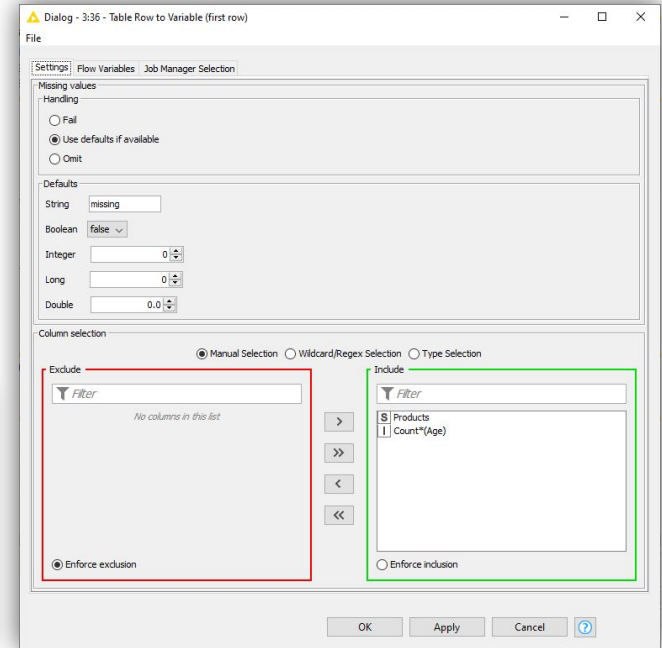
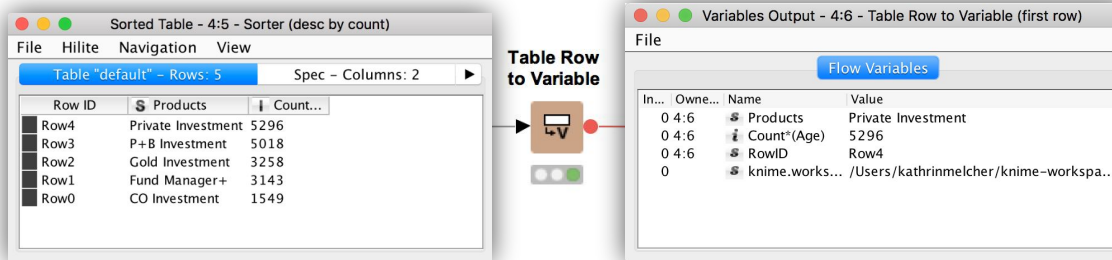
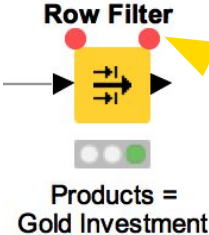


Table Row to Variable

- Takes a table as input and converts the first row to Flow Variables
 - Column names -> Flow Variable names
 - Column values -> Flow Variable values
- Only the first row is transformed, additional rows are discarded

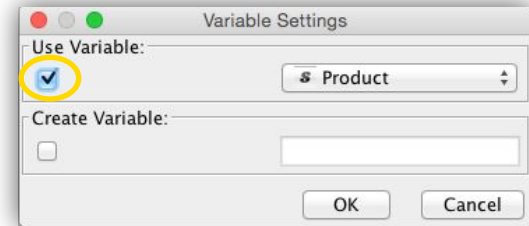
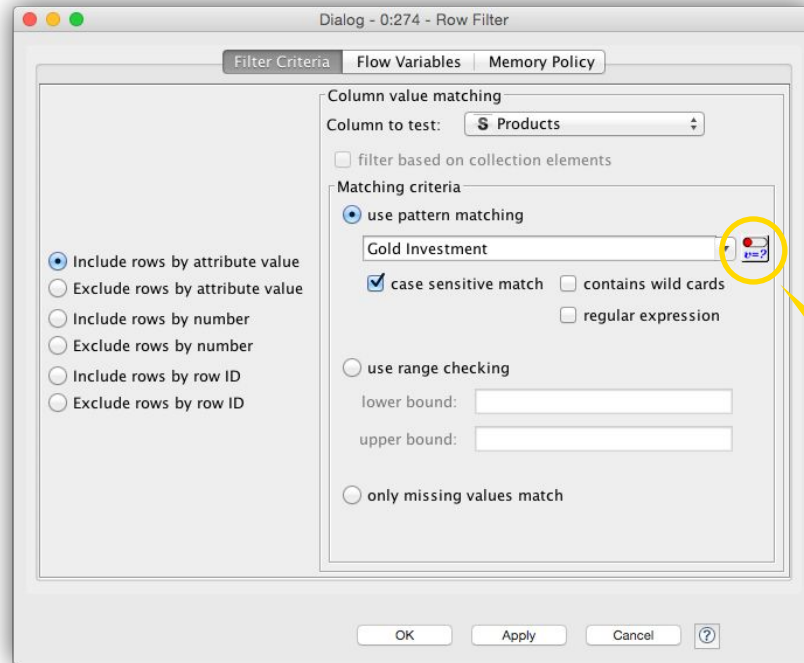


Flow Variable Ports



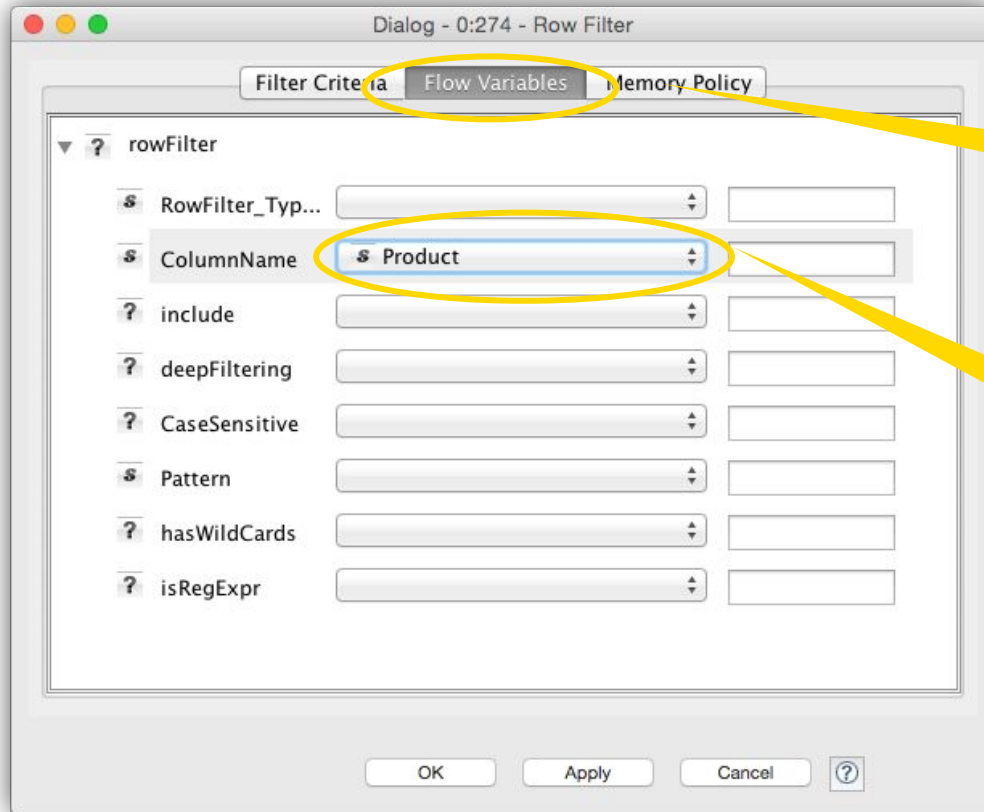
	Configure...	F6
	Execute	F7
	Execute and Open Views	Shift+F10
	Cancel	F9
	Reset	F8
	Edit Node Description...	Alt+F2
	New Workflow Annotation	
	Connect selected nodes	Ctrl+L
	Disconnect selected nodes	Ctrl+Shift+L
	Create Metanode...	
	Create Component...	
	Compare Nodes	
	Show Flow Variable Ports	
	Cut	
	Copy	
	Paste	
	Undo	
	Redo	
	Delete	
	Filtered	

Apply a Flow Variable (Button)



The Flow Variable button

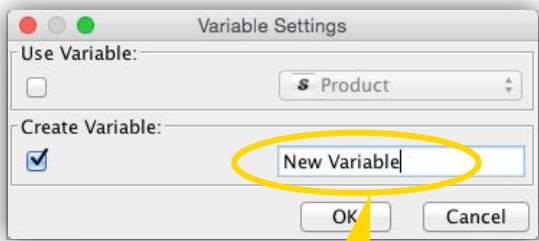
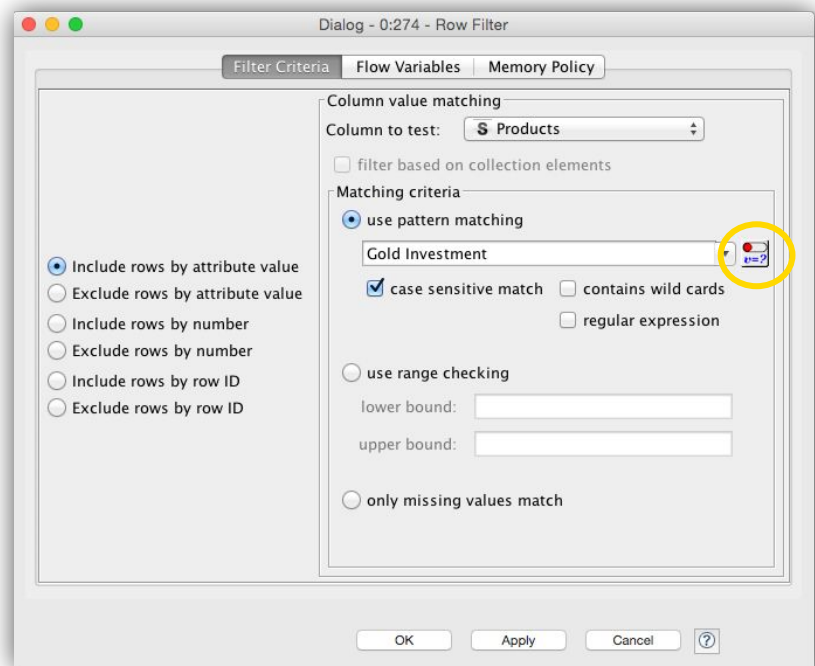
Apply a Flow Variable (Advanced)



The Flow Variables
tab

List of available
Flow Variables

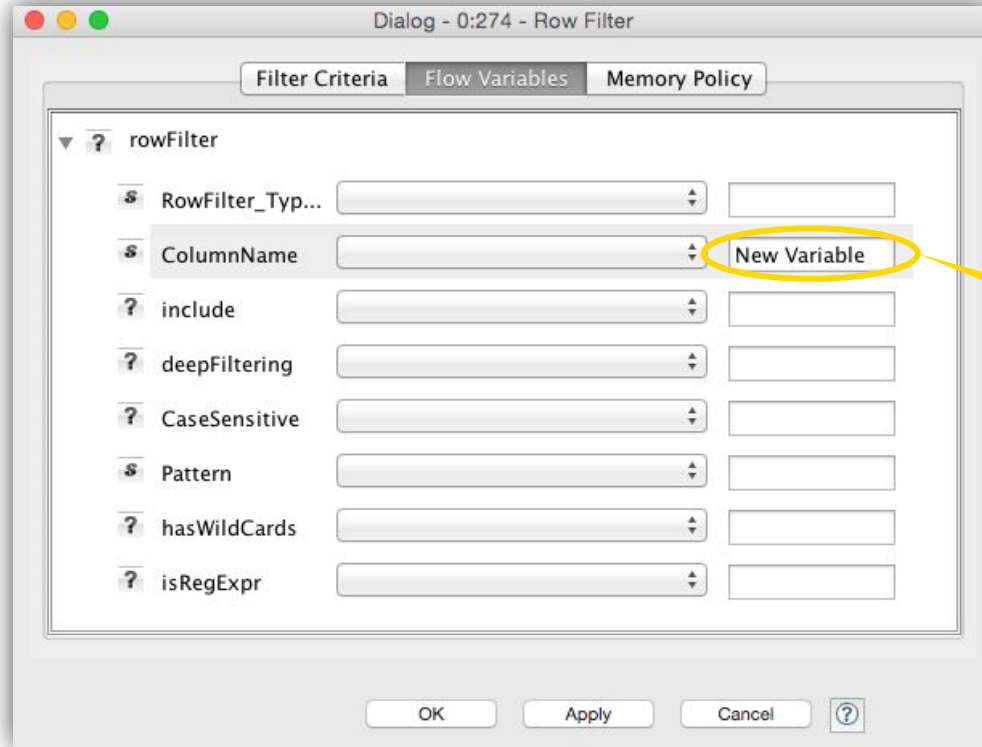
Create a Flow Variable (Button)



Name of the new
Flow Variable

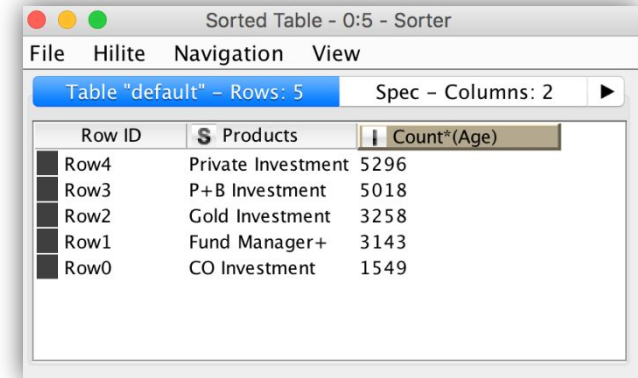
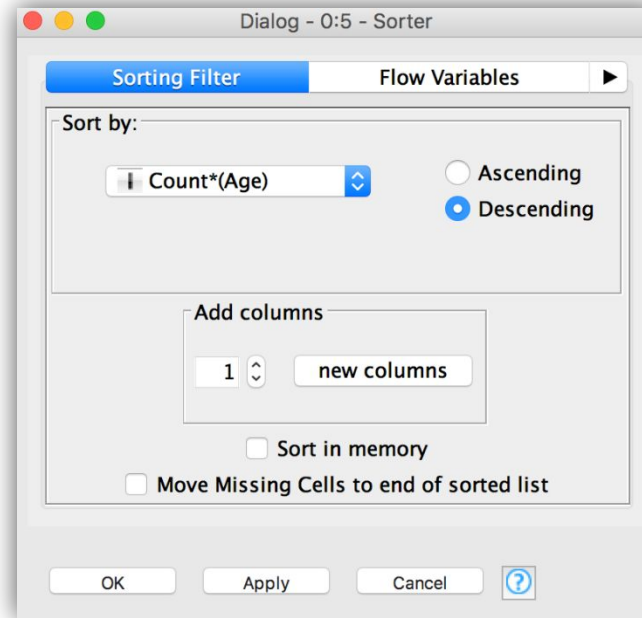
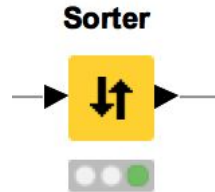
Create a Flow Variable (Advanced)

- Converting a setting value into a Flow Variable



Sorter

- Sorts a table!
- Choice of ascending or descending
- Sort by multiple columns

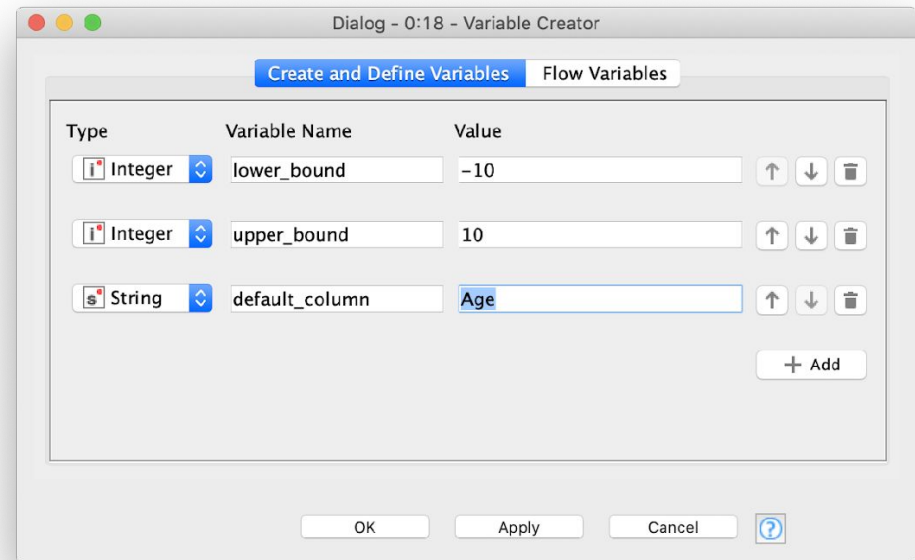
A screenshot of the "Sorted Table - 0:5 - Sorter" window. It has a menu bar with "File", "Hilite", "Navigation", and "View". Below the menu bar is a status bar showing "Table 'default' - Rows: 5" and "Spec - Columns: 2". The main area displays a table with three columns: "Row ID", "Products", and "Count*(Age)". The rows are sorted by "Count*(Age)" in descending order.

Row ID	Products	Count*(Age)
Row4	Private Investment	5296
Row3	P+B Investment	5018
Row2	Gold Investment	3258
Row1	Fund Manager+	3143
Row0	CO Investment	1549

Variable Creator

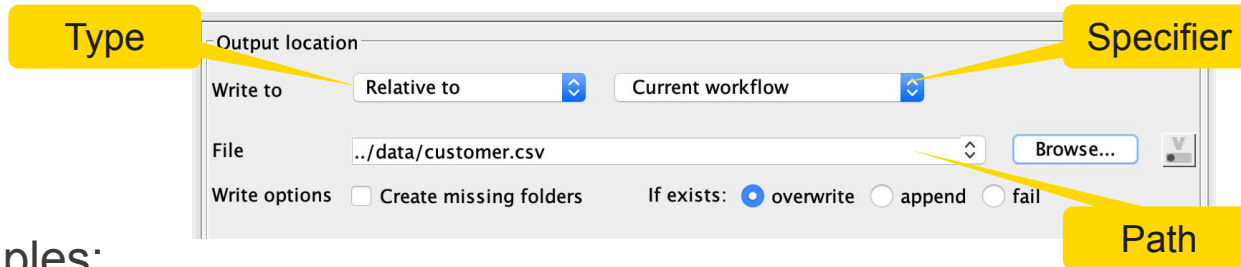
- Allows to create flow variables of different types
- Click on “+ Add” to add a new variable and define a custom
 - Variable Name
 - Variable Value

Variable Creator



Path Variables

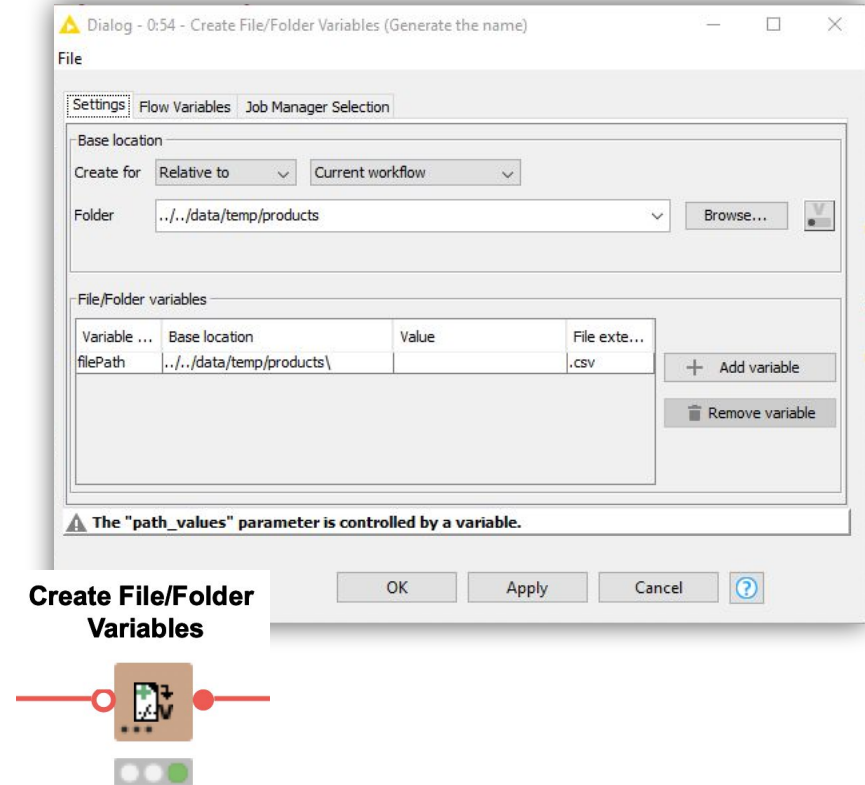
- Special flow variable type to point to a file or folder
 - E.g. to control output location of a file
- A path type consists of three parts:
 - **Type**: Specifies the file system type e.g. local, relative, mountpoint, custome_url or connected.
 - **Specifier**: Optional string with additional file system specific information e.g. relative to which location (knime.workflow)
 - **Path**: Specifies the location within the file system



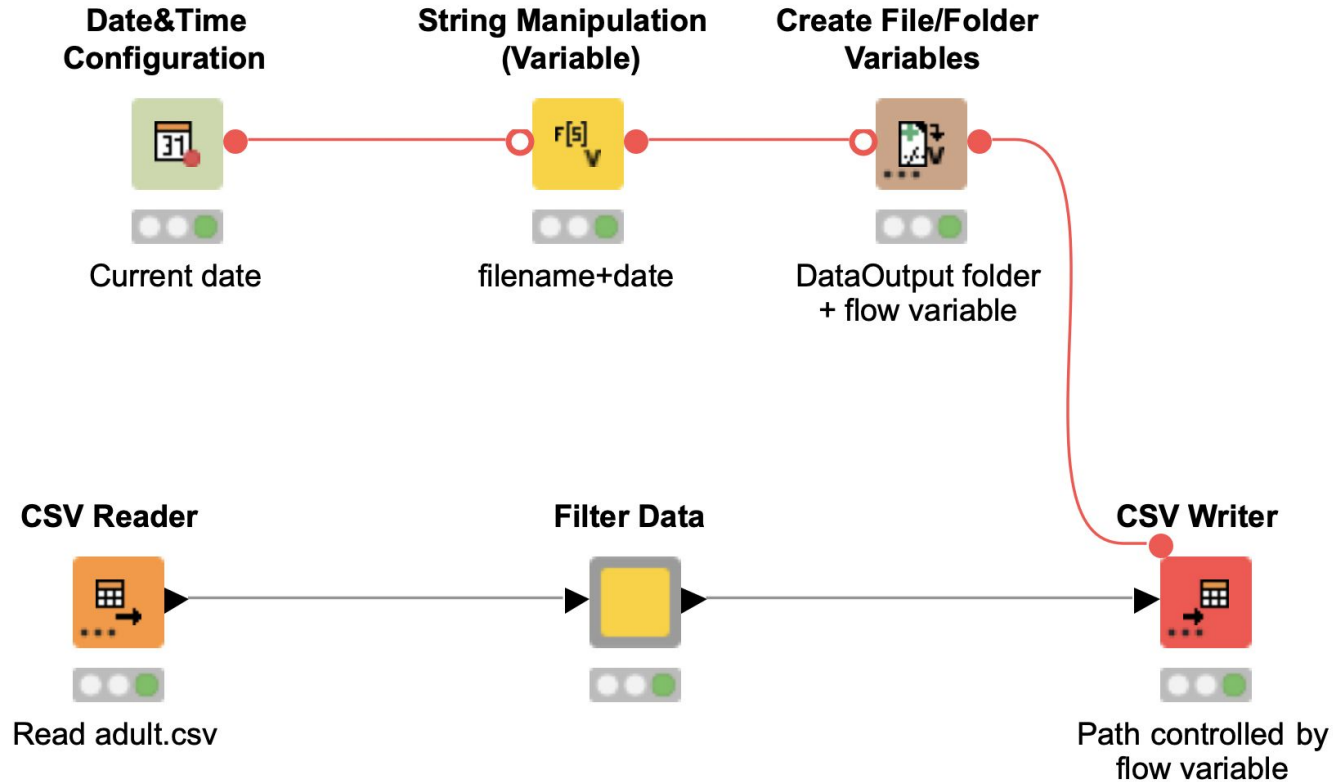
- Examples:
 - (LOCAL, , C:\Users\username\Desktop)
 - (RELATIVE, knime.workflow, file1.csv)
 - (MOUNTPOINT, MOUNTPOINT_NAME, /path/to/file1.csv)
 - (CONNECTED, amazon-s3:eu-west-1, /mybucket/file1.csv)

Create File/Folder Variables

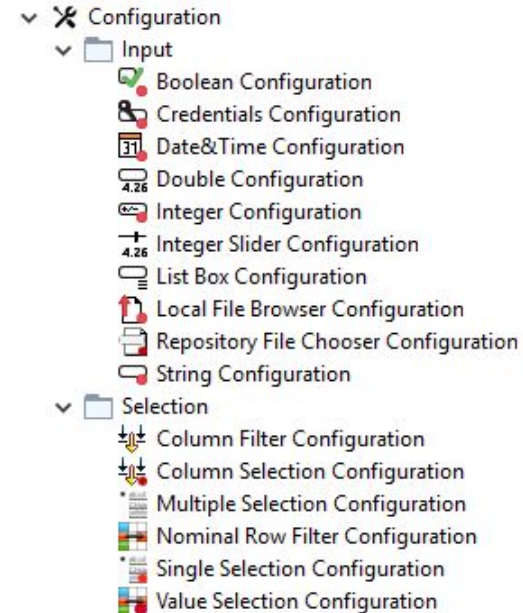
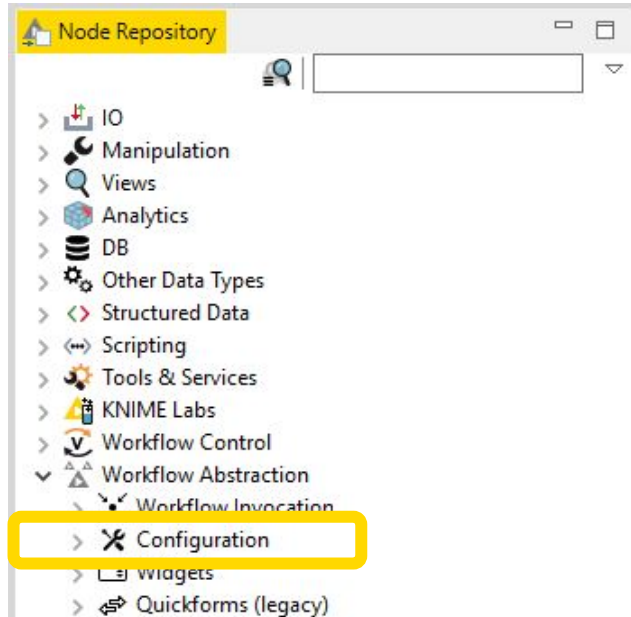
- Creates one or multiple path flow variable(s) pointing to files / folders
- Inputs:
 - Base location
 - Flow variable name(s)
 - Value (file name or path relative to base location)
 - File extension (optional)
- Output variables can be used to control the output location in writer nodes.



Example: Add Execution Date to File Name

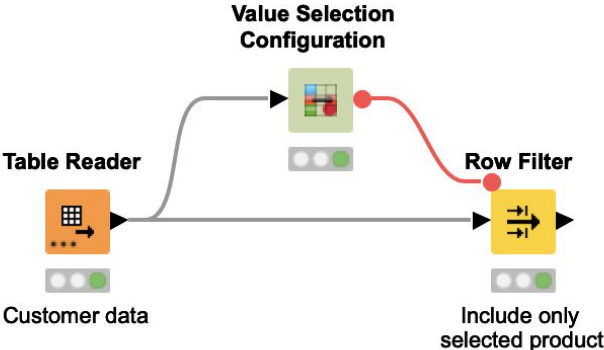


Configuration Nodes for Variable Creation and Output



Configuration Node Configuration

Use Configuration nodes to create Flow Variables



Dialog - 2:14 - Value Selection Configuration

Control | Flow Variables | Job Manager Selection | Memory Policy

Label: Select product:

Description: Select the product, which should be included.

Parameter/Variable Name: product_select

Selection Type: Dropdown

Lock Column: ☒

Default Column: \$ Products

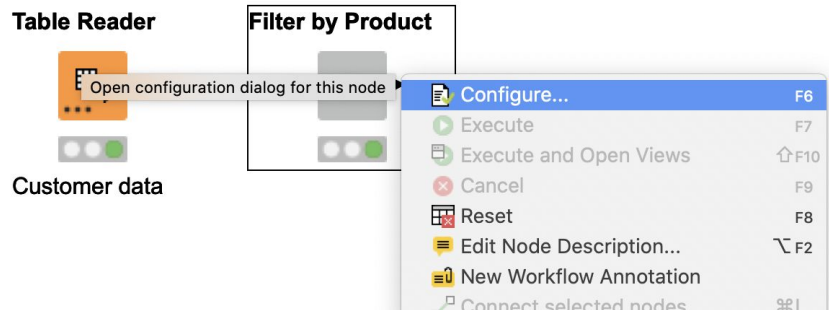
Default Value: Gold Investment

Limit number of visible options: ☐

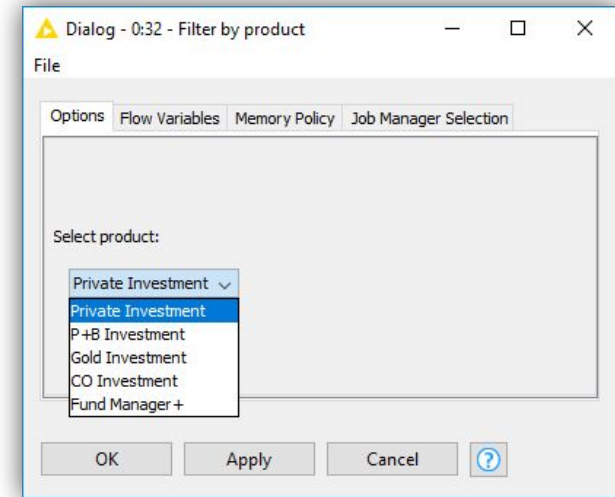
Number of visible options: 10

OK Apply Cancel ?

Simple Configuration of Component



- Double click a component to configure it
- For use on the KNIME Business Hub as a Data App, replace Configuration nodes with Widget nodes



Multiple Selection Configuration Node

Dialog - 0:57:0:55 - Multiple Selection Configuration

Control Flow Variables Memory Policy

Label:

Description:

Parameter/Variable Name:

Selection Type:

Possible Choices:

Default Values:

Limit number of visible options: ☐

Number of visible options:

OK Apply Cancel ?

Multiple Selection
Configuration



Dialog - 0:51 - Multiple Selection Configuration

Control Flow Variables Memory Policy

defaultValue

hideInDialog

parameterName

label

description

required

flowVariableName

possible_choices

type

limit_number_visible_options

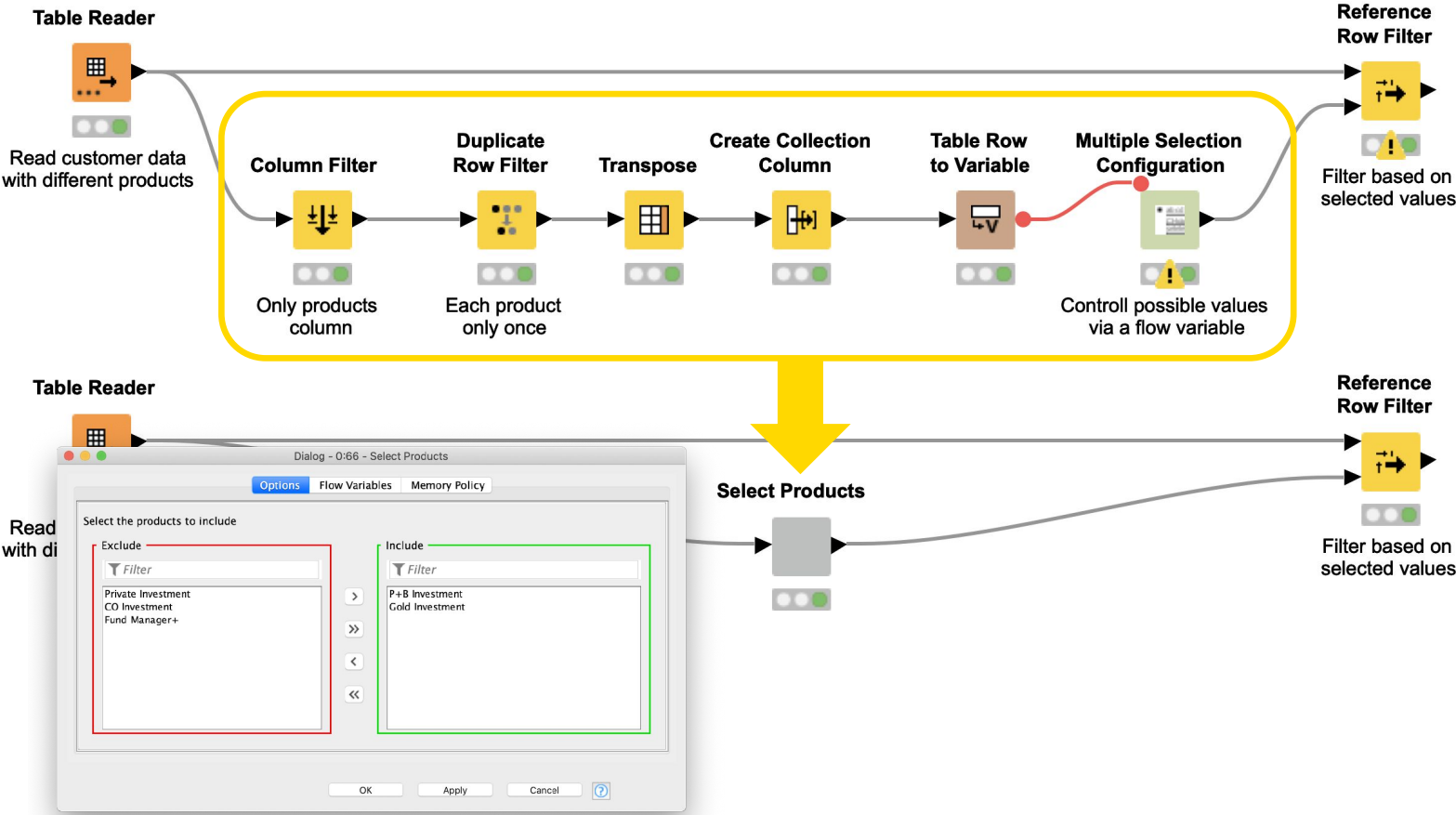
number_visible_options

List of Possible Choices can
be controlled by an array flow
variable

Array Variables

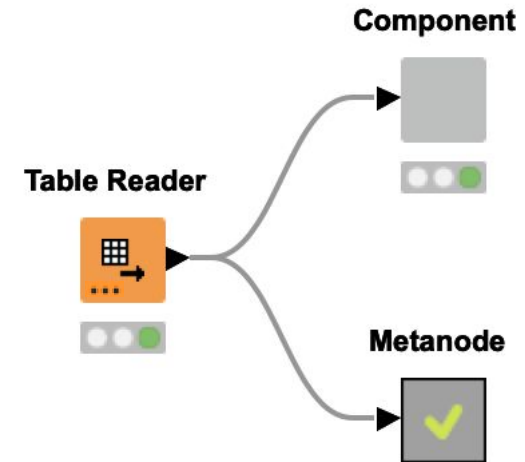
- An array flow variable is a flow variable with many values
- Can be used to control setting options with more than one value, e.g.
 - Include / exclude of columns
 - Define list of possible values
 - Control the variable names and values in the Create File / Folder Variables node
- How to create an array flow variable?
 - Step 1: Create a collection cell (e.g. Create Collection Column node or GroupBy node)
 - Step 2: Use the Table Row to Variable node to create the array flow variable based on the collection cell

Example: Creating an Array Flow Variable

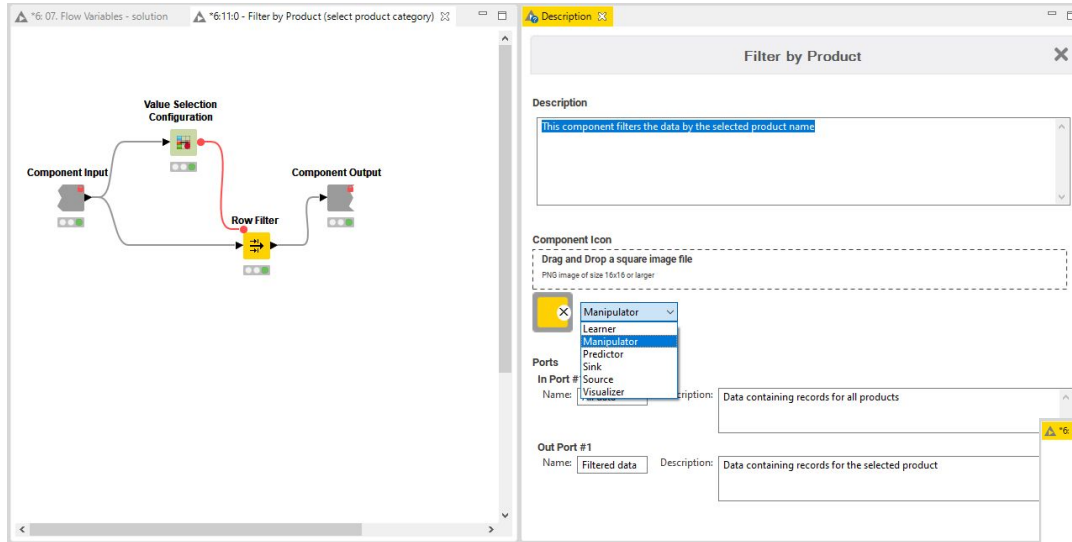


Components

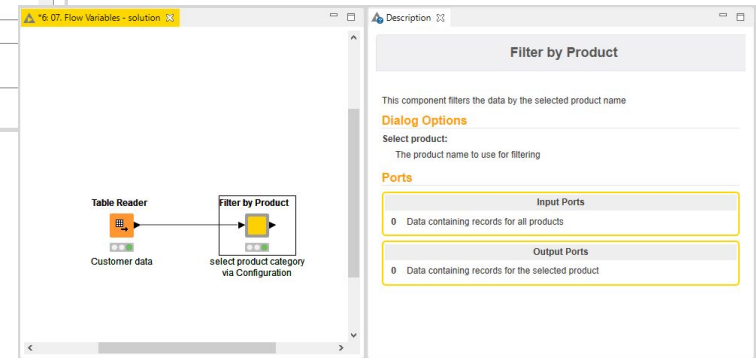
- Encapsulate a functionality for reuse and sharing
- Main features:
 - Local Flow Variable scope
 - Configurable via Configuration nodes
- Key to advanced functionality in KNIME products
 - Component corresponds to a KNIME Data App page
 - Configurations on a Data App page are defined using Widgets
 - Can be shared via KNIME Community Hub



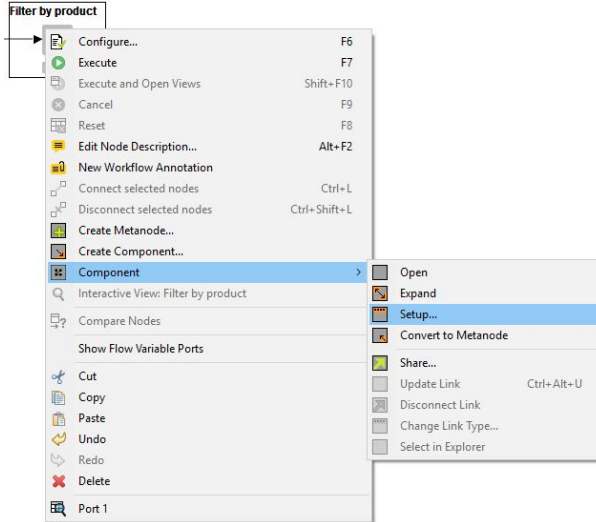
Component Description



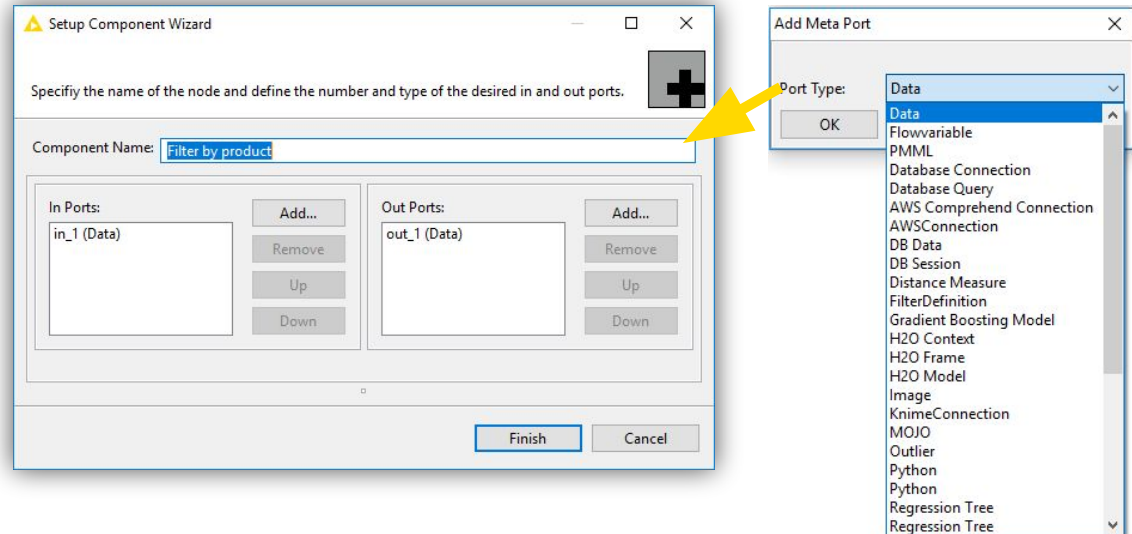
- Double click a component to configure it
- For use on the KNIME Business Hub as Data App, replace Configuration nodes with Widget nodes



Configure Component Ports

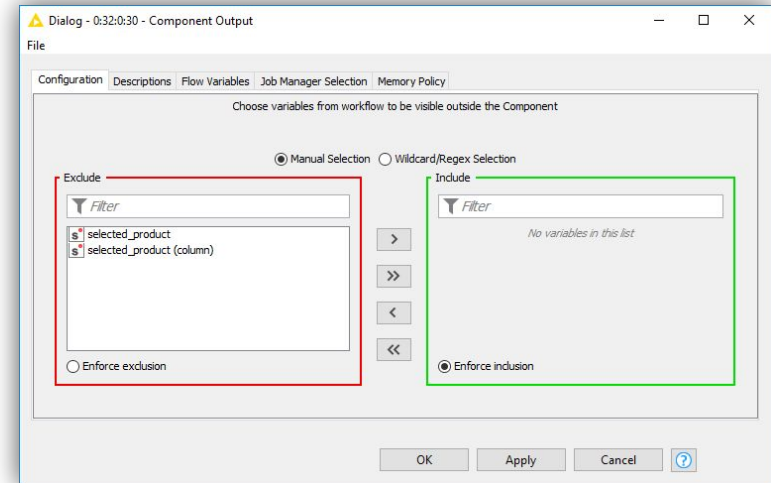
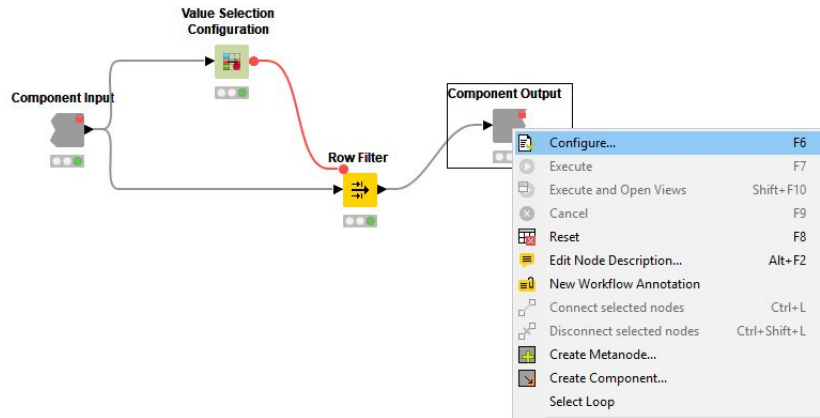


- Add input and output ports to metanodes/components
- Remove ports to adapt to changes after creation of the metanode/component



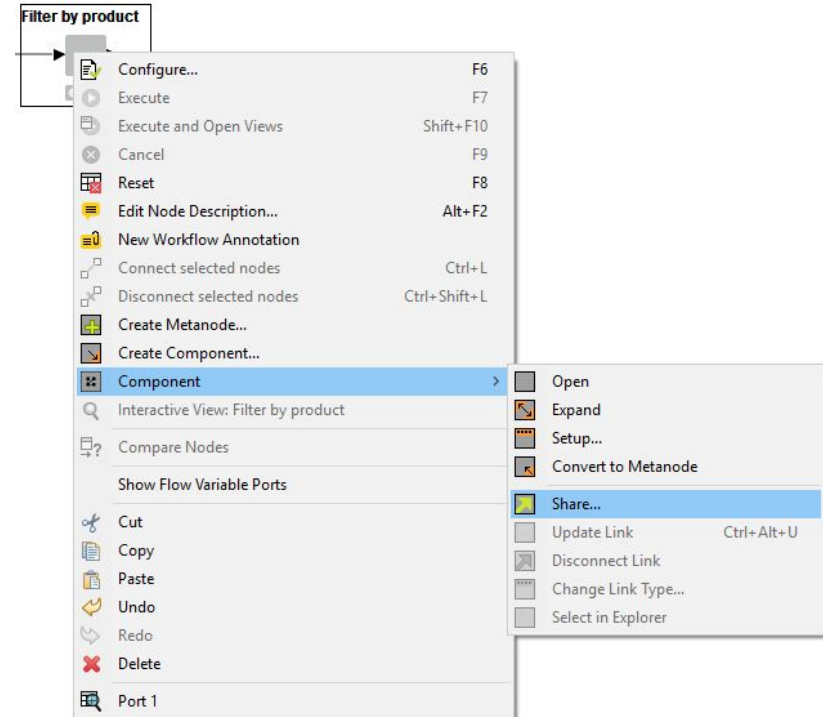
Passing Variables from Components

- Flow Variables are -by default - only available locally inside the component
- Configure the component input/output to pass Flow Variables from/to outside the component



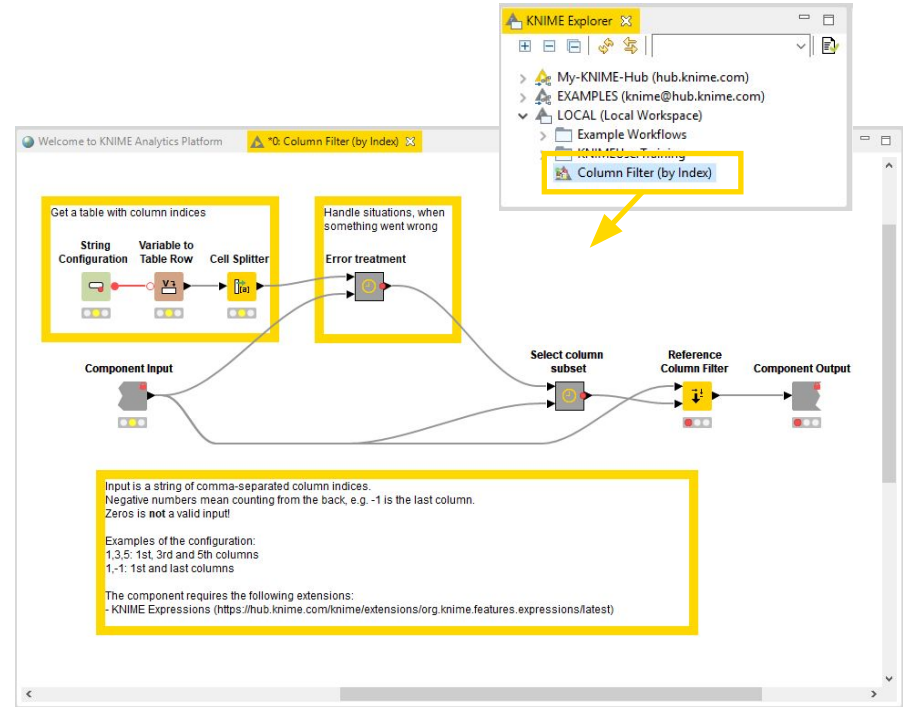
What is a Shared Component?

- Components can be saved in your KNIME workspace for later reuse
- To do this, simply right-click any component and select “Share...”
- Shared components are read-only instances of a component
- Public Shared Components are available on the EXAMPLES Server and on the KNIME Community Hub



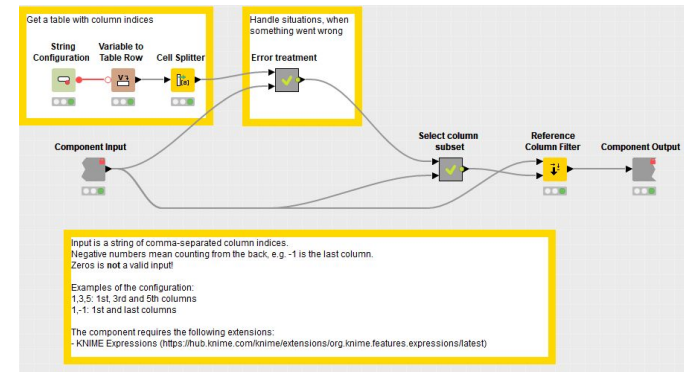
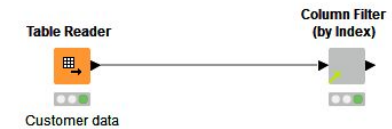
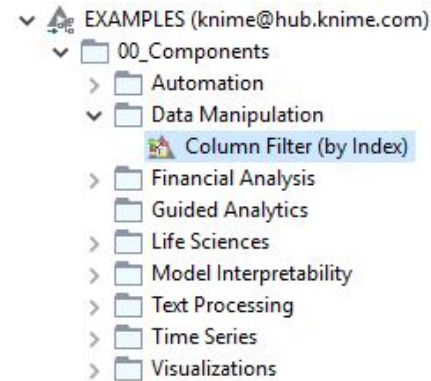
How can you Edit a Shared Component?

- Components can be edited using the Component Editor, similar to workflows
- To edit a component using the Component Editor, double-click the component in its location in the KNIME Explorer
- To ensure components are executable when opened in the Component Editor, choose the option to “Include input data with component” when sharing it

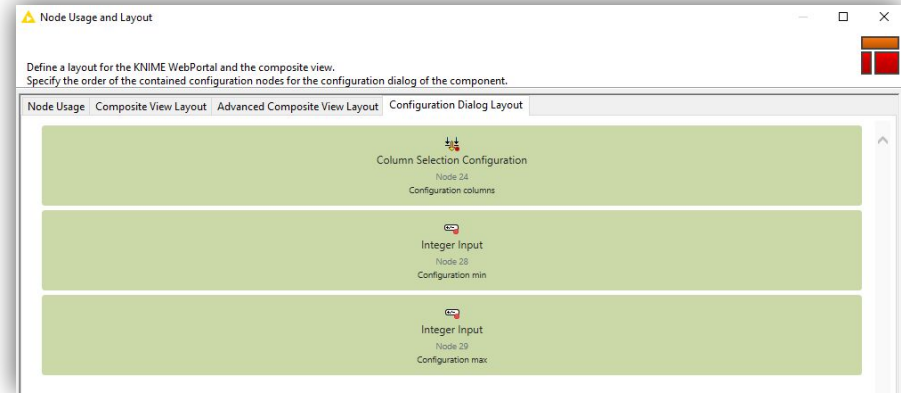
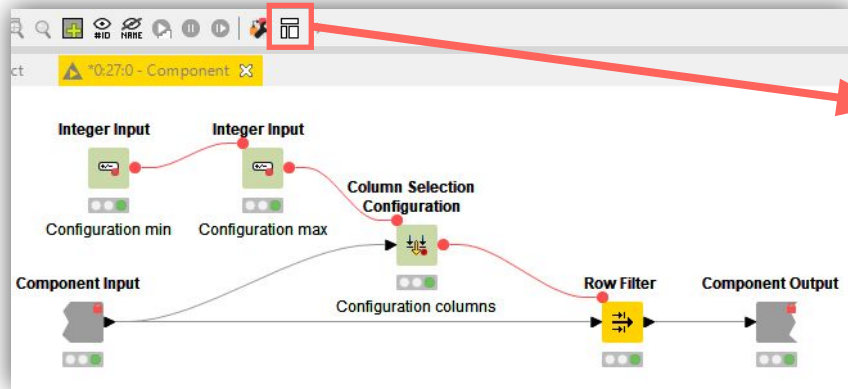


How can you Use a Shared Component?

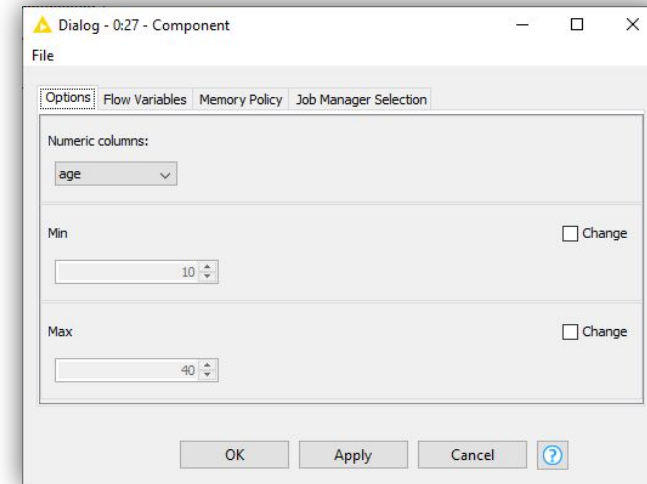
- To use a Shared Component, drag and drop it to your workflow editor
- Instances of Shared Components can be updated either manually or when the workflow is opened
- A Shared Component can also be unlinked from its original location, which makes it editable in the workflow directly
- Update Shared Components by overwriting them



Configuration Dialog Layout

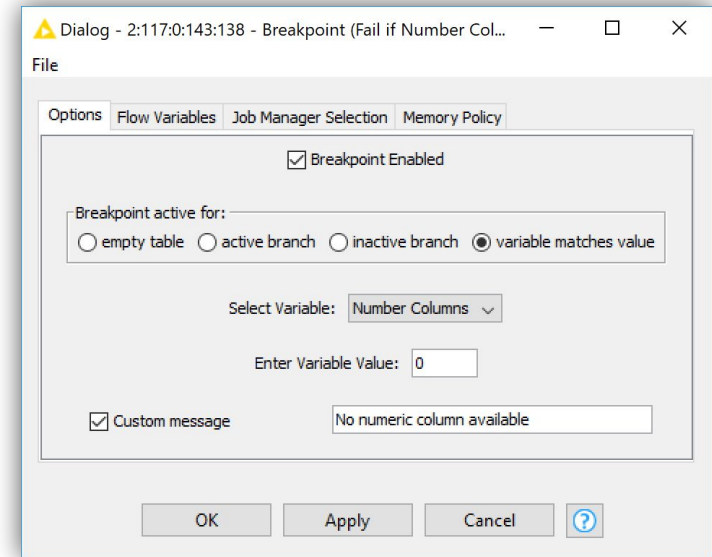
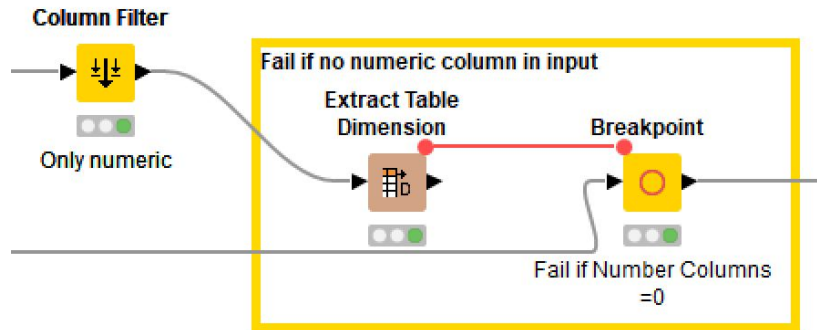


- Click layout button when inside component to modify the order of the setting options in configuration window of the component

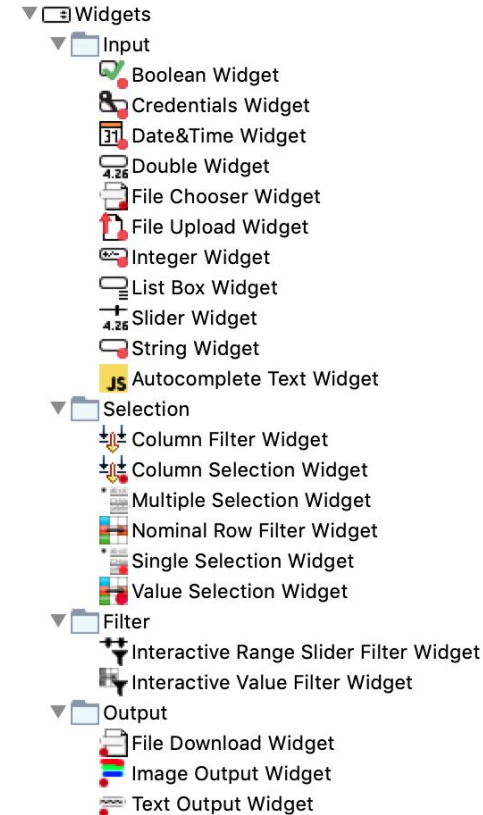
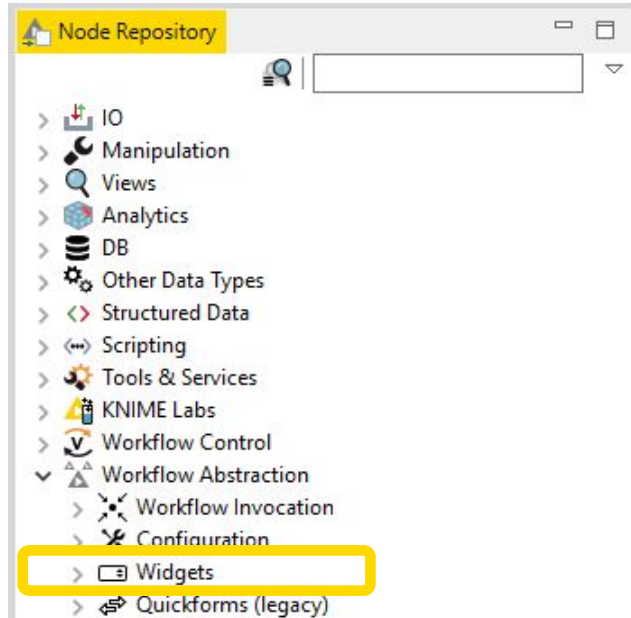


Breakpoint

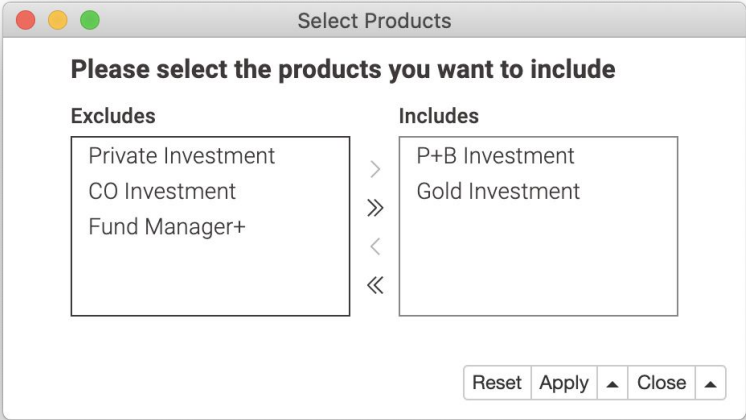
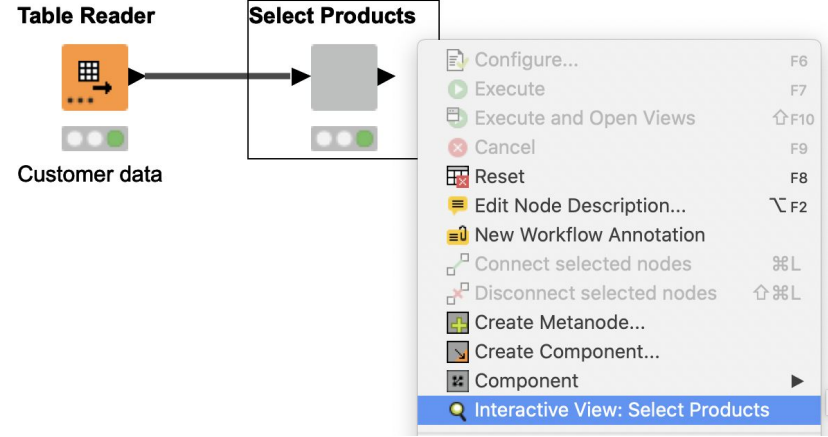
- Stops execution of a workflow branch
- Useful to stop the execution of a component and provide a custom error message
- Execution stops based on the selected condition:
 - Empty table
 - Active/Inactive branch
 - Flow Variable value



Widget Nodes for Variable Creation and Output



Simple Configuration of Component



Summary: Flow Variables

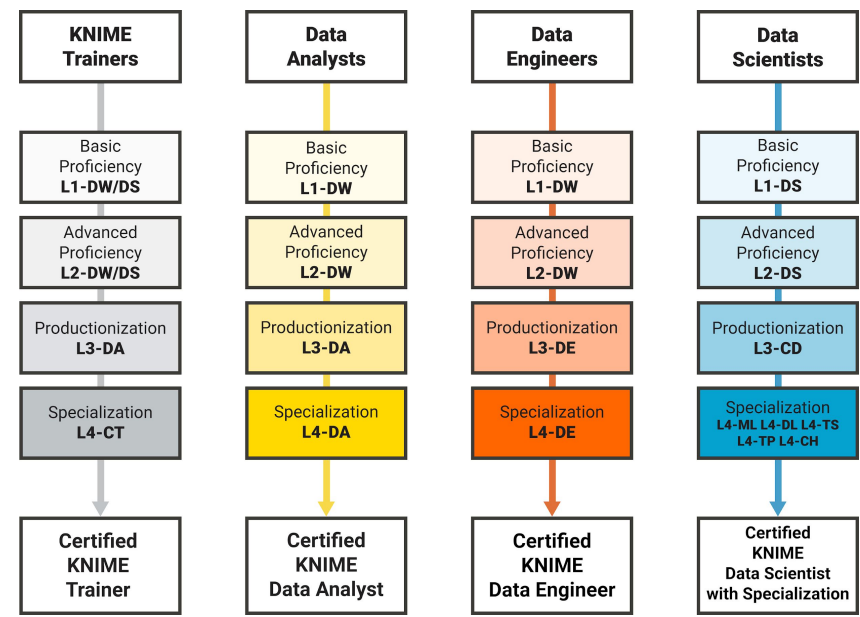
- Flow Variables are workflow parameters used to overwrite existing node settings
- A Flow Variable is carried along workflow branches (parallel branches don't share local Flow Variables)
- Flow Variables can be of type String, Integer, Double, Boolean, Long, Paths and Array
- Flow Variables can be created
 - 1. in the “Flow Variables” tab of any node
 - 2. using specific nodes, e.g. Table Row to Variable, Variable Creator
 - 3. using Configuration and Widget nodes

Exercise: 10_Flow_Variables

- Activity I: Filter the customer data to
 - Customers of the "Gold Investment" product
 - Customers of the most common product in the data
- Activity II: Create a component that allows a user to choose an investment product and filter the data by that product
- Activity III (optional): Create a path variable that automatically has the current execution date in the file name and write the filtered table into a CSV file
- Activity IV (optional): Create a component that allows to select multiple products out of all available products, using a flow variable of type array

KNIME Learning Paths

From level **L1** to level **L4** for various professional profiles



Self-paced courses: videos and exercises at your own pace and for free

Instructor led courses: Scheduled sessions and guided exercises in paid courses



knime.com/knime-courses

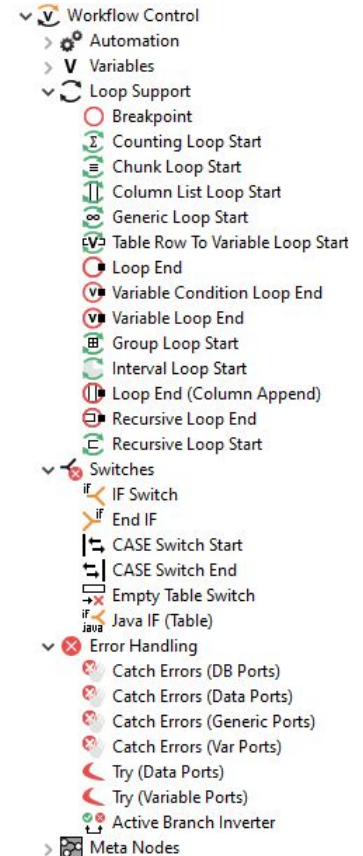
Workflow Control

Loops, Switches, Try-Catch



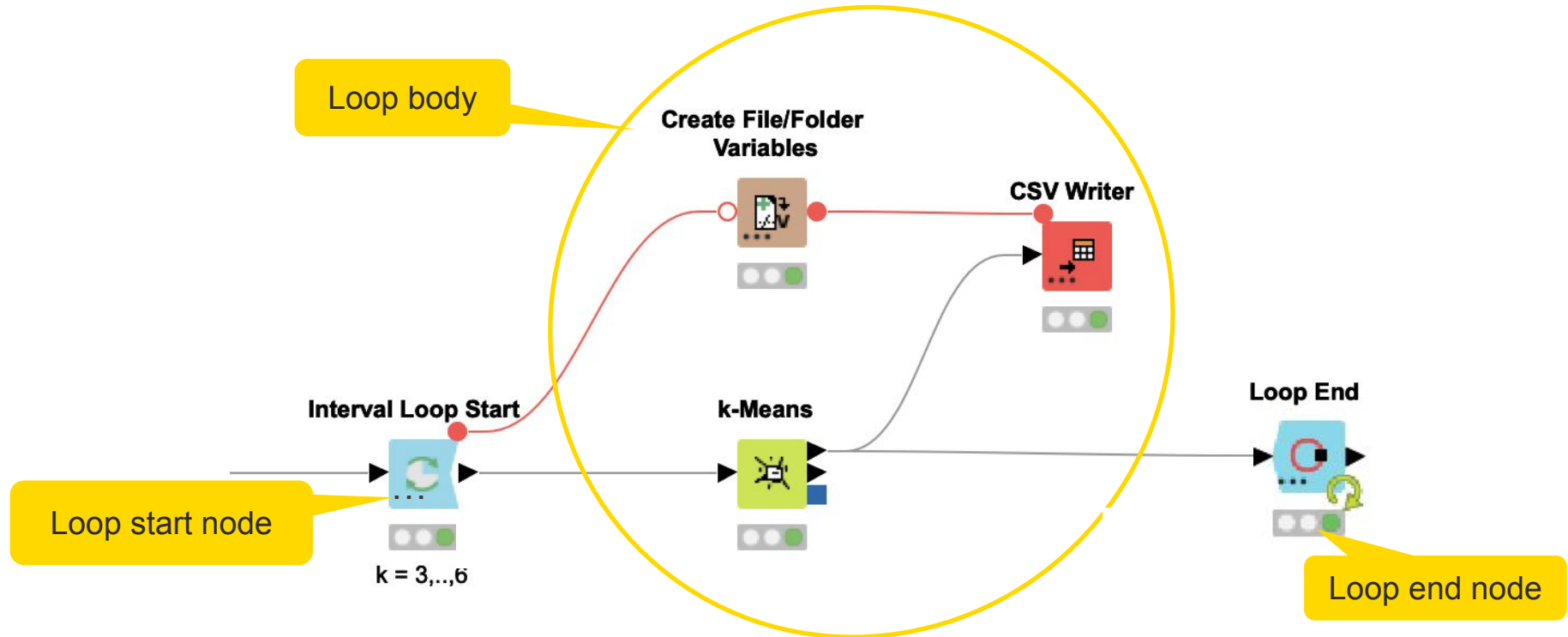
Workflow Control Structures

- **Loops**
 - Iterate over a workflow snippet with variable inputs.
- **Switches**
 - Direct the path of a workflow by selectively executing one or more workflow branches.
- **Try-Catch**
 - Handle workflow branches that may fail in execution - when you don't know about this before executing



The Loop Block

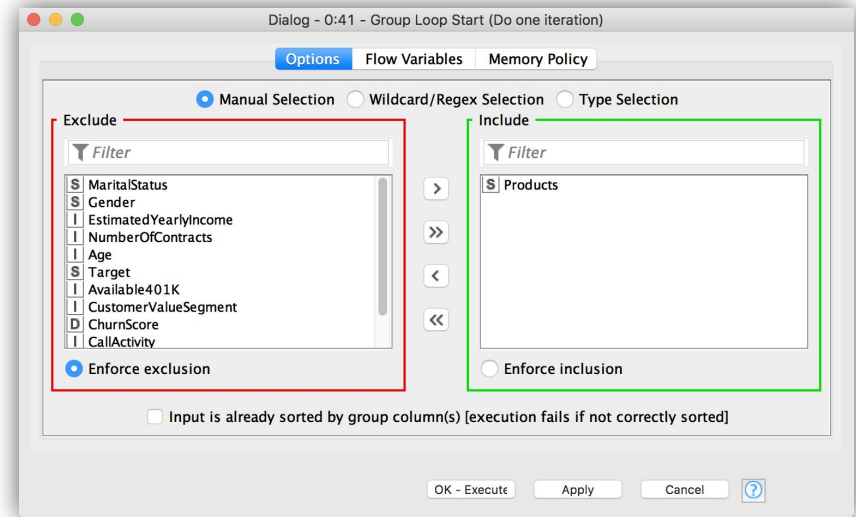
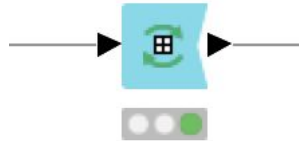
- A loop block is defined by the appropriate loop start and loop end nodes.
- Loop body = the nodes in between (including the side branches).



Group Loop Start

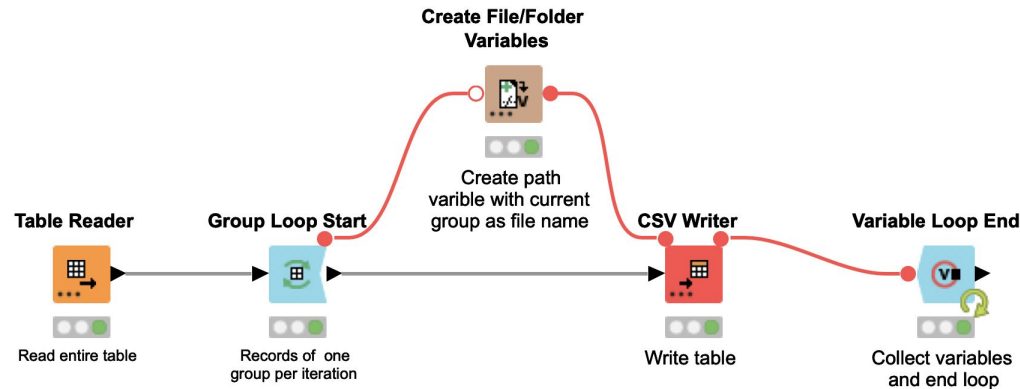
- Similar to GroupBy except without aggregation tab.
- Each iteration of the loop passes the next group of rows.
- You can implement an aggregation task. It can be anything from a complex calculation to updating a database.

Group Loop Start



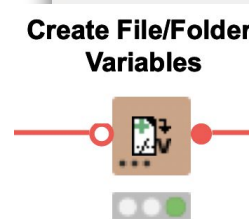
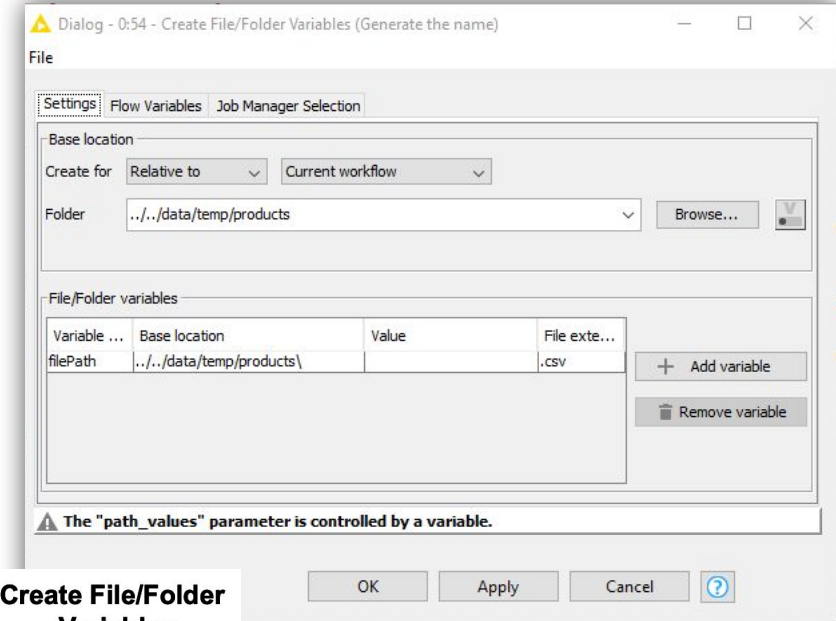
Example: Writing Aggregated Files

- Group Loop Start □ Variable Loop End
- Group data by specific column values
- Iterate over all groups of data
- Create an appropriate file name
- Write grouped data to tables with new file name

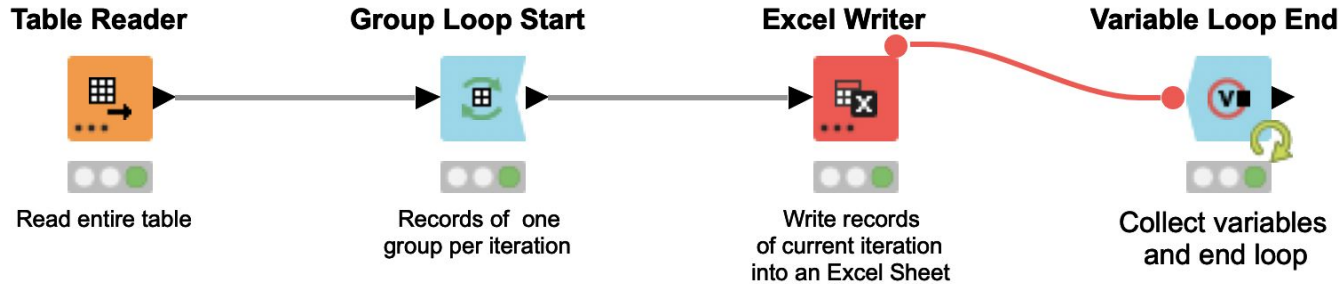


Create File/Folder Variables

- Creates one or multiple path flow variable(s) pointing to files / folders
- Inputs:
 - Base location
 - Flow variable name(s)
 - Value (file name or path relative to base location)
 - File extension (optional)
- Output variables can be used to control the output location in writer nodes.



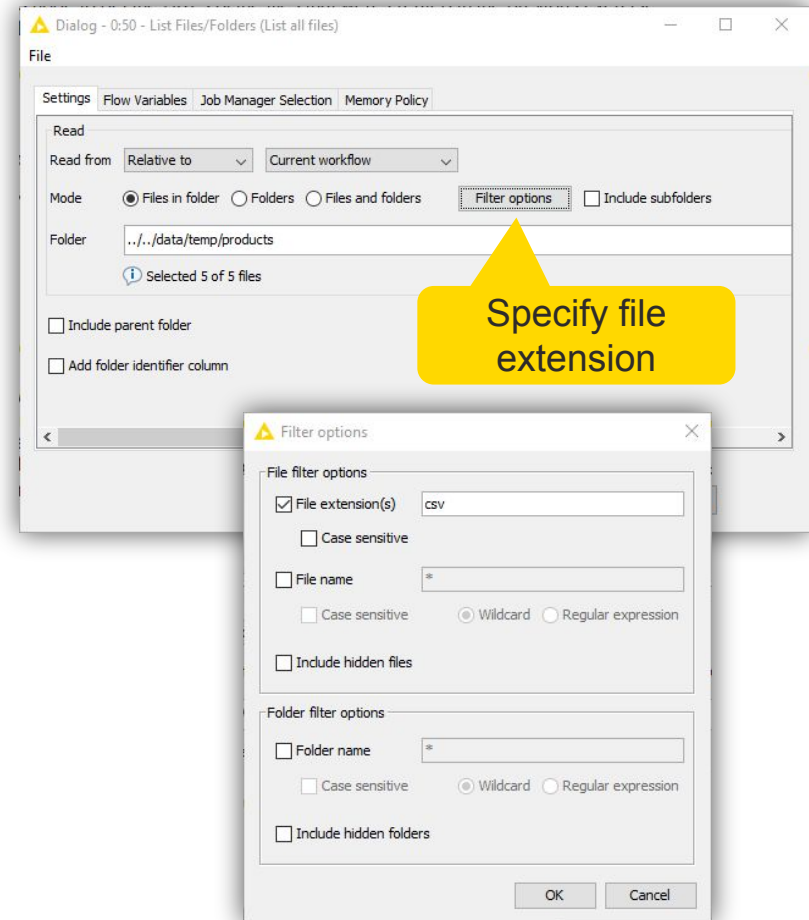
Example: Writing Multiple Excel Sheets



List Files/Folders

- List all files in a directory
- Restrict to:
 - Top level directory (i.e. not recursive)
 - Specific file extensions
 - Matching name patterns (regex or wildcard)
- Provide file references as a table of URLs and absolute paths

List Files/Folders



Example: Reading Many Excel Sheets

- List all sheet names of an Excel file
- Convert sheet name into a flow variable (1 sheet name per iteration)
- In each iteration, read the spreadsheet with the current sheet name
- Close the loop and collect the results

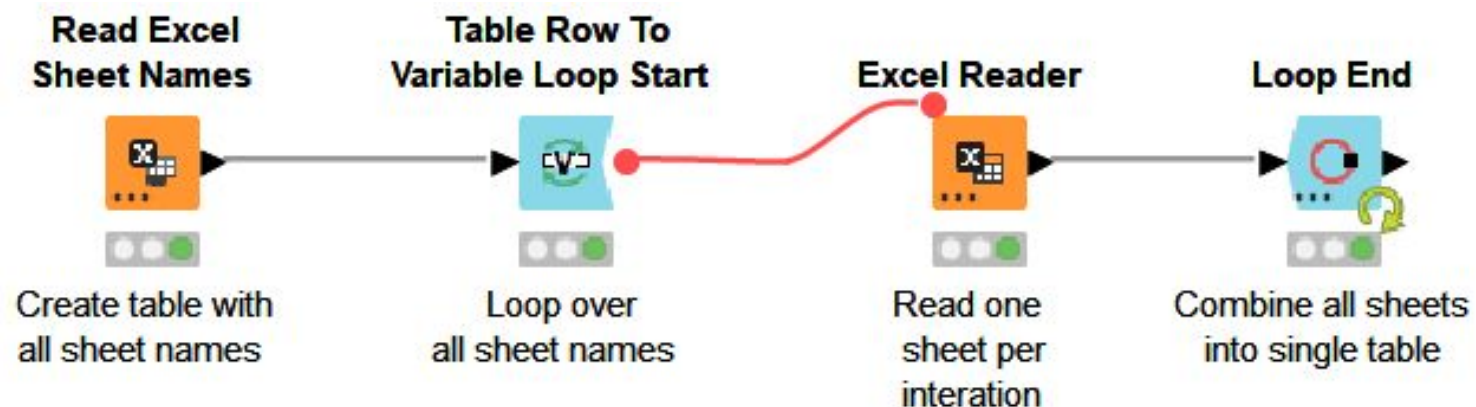
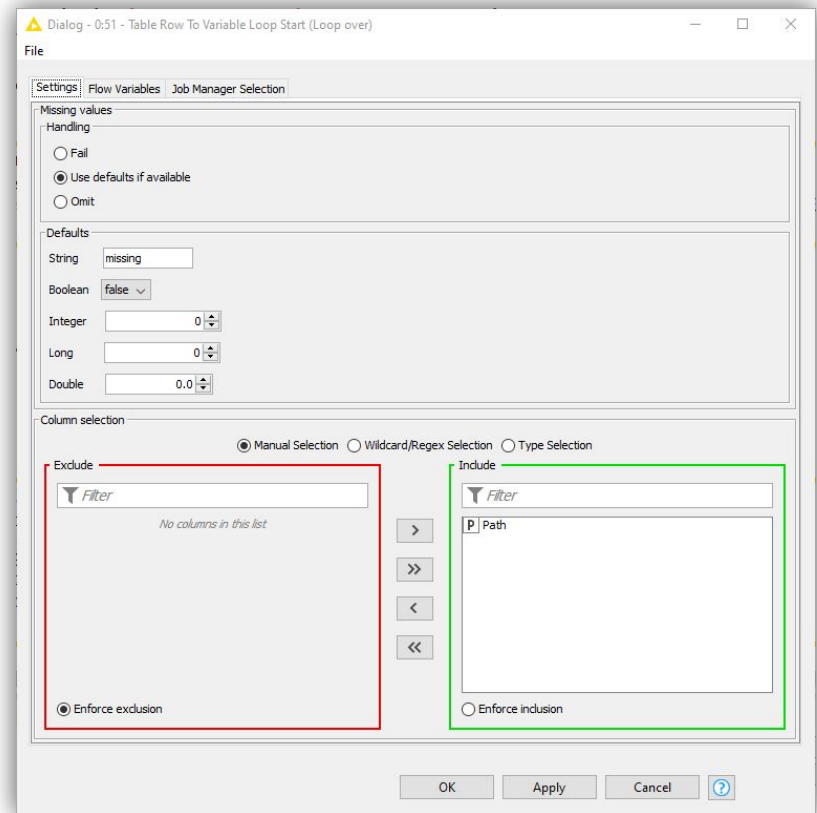
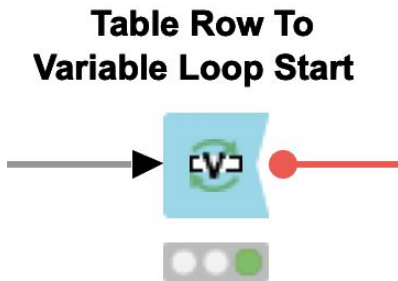


Table Row to Variable Loop Start

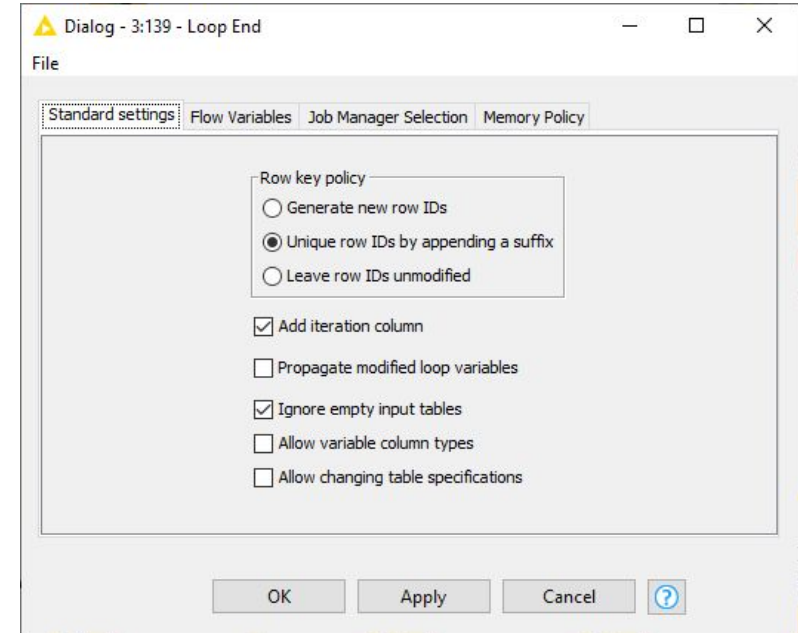
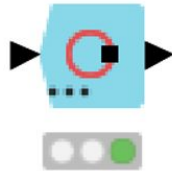
- Similar to the Table Row to Variable node
- Each iteration of the loop converts the next row of the input table into Flow Variables
- Injects variables into other nodes to re-execute subflows with a progression of settings



Loop End

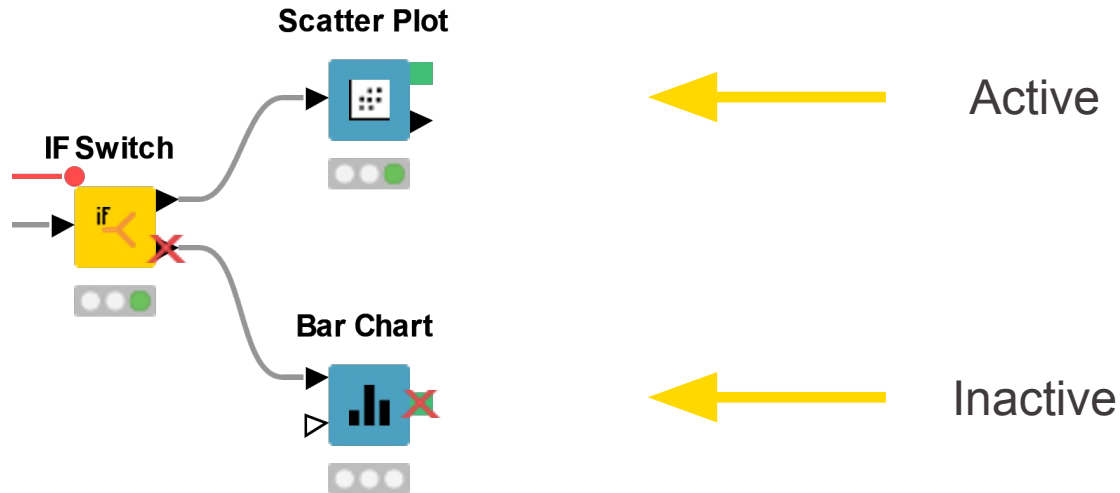
- Can be used to end of a loop
- Collects the results of the different iterations by row-wise concatenation of the incoming tables
- Provides options to:
 - Add a column with the iteration number
 - Propagate modified loop variables
 - Allow variable column types
 - Allow changing table specifications

Loop End



Switches

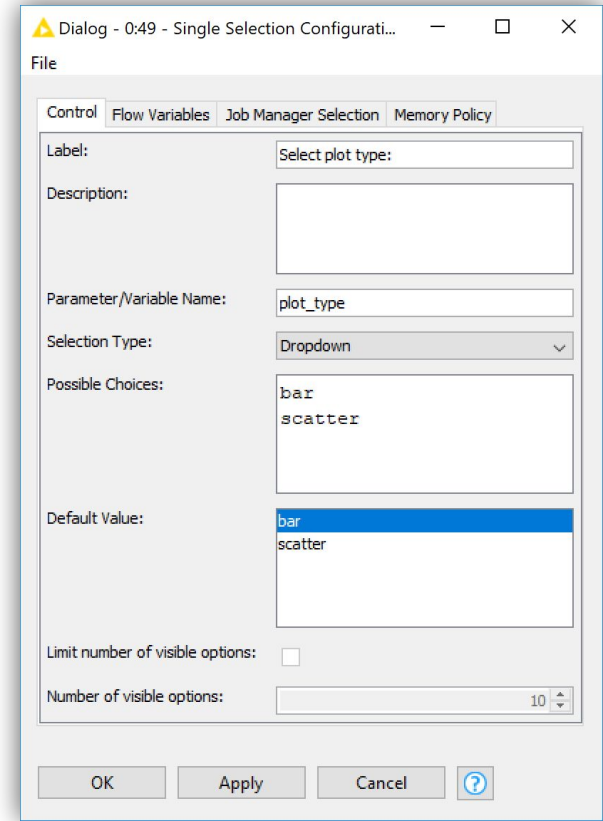
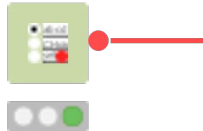
- A switch allows you to selectively activate branches of a workflow
- Inactive branches are marked with a red x on their output ports. Inactive nodes propagate down stream.



Single Selection Configuration

- Configuration: Select single value from list of Strings
- Returns selection as string type Flow Variable
- Choose between different layout options (dropdown, radio buttons...)

Single Selection Configuration



Dialog - 0:49 - Single Selection Configuration

File

Control Flow Variables Job Manager Selection Memory Policy

Label: Select plot type:

Description:

Parameter/Variable Name: plot_type

Selection Type: Dropdown

Possible Choices: bar
scatter

Default Value: bar
scatter

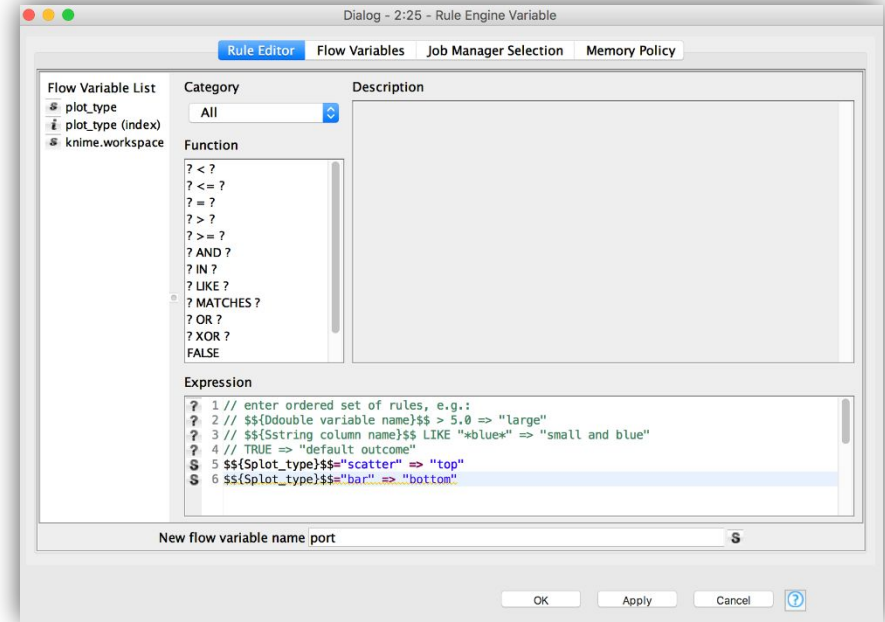
Limit number of visible options: ☐

Number of visible options: 10

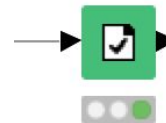
OK Apply Cancel ?

Rule Engine/Rule Engine Variable

- Defines custom logic for using simple rules.
- Rules like: **<Antecedent>** => **<Consequence>**
 - (1=1 => "true")
- May be used in Flow Variables or tables
- Easiest way to encode logic for switches



Rule Engine

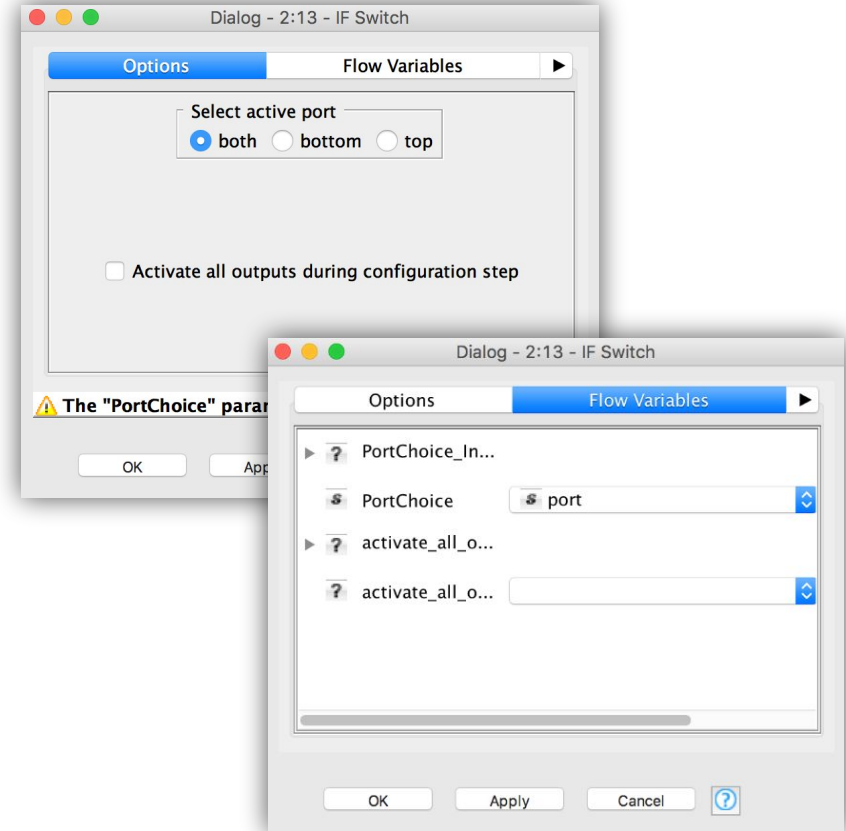
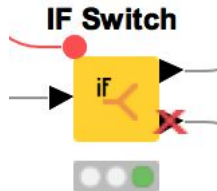


Rule Engine Variable



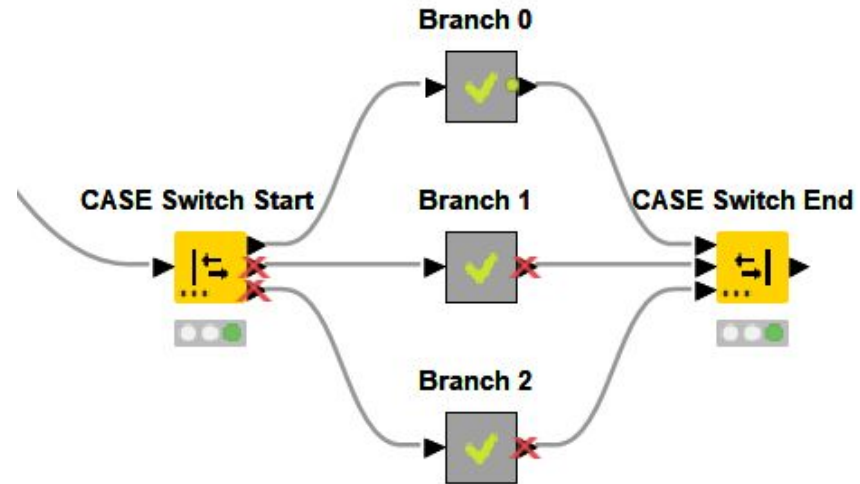
If Switch

- Controls which branches of your workflow are active programmatically
- Controlled with a Flow Variable, setting the value to the literal Strings: “top”, “bottom”, “both”
- May be used in Flow Variables or tables (different nodes)



Case Switch Start & End

- Similar to If-Switch: Takes data from single input port and passes it to the active output port
- Nodes connected to inactive branches are not executed



Case Switch Start & End

■ Case Switch Start

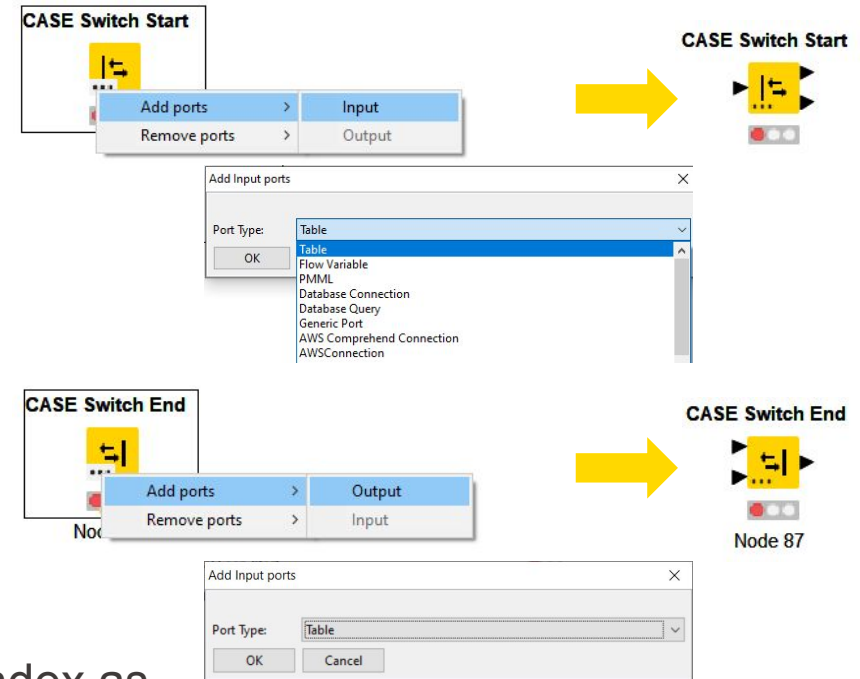
- Add an input port with a specific type (e.g., Data)
- ☐ Two output ports are also added
- ☐ Additional output ports can be added

■ Case Switch End

- Add an output port with a specific type (e.g., Data)
- ☐ Two input ports are also added
- ☐ Additional input ports can be added

■ Configure via node dialog, or pass port index as Flow Variable

- From the top, 0, 1, 2, ... (however many ports there are)



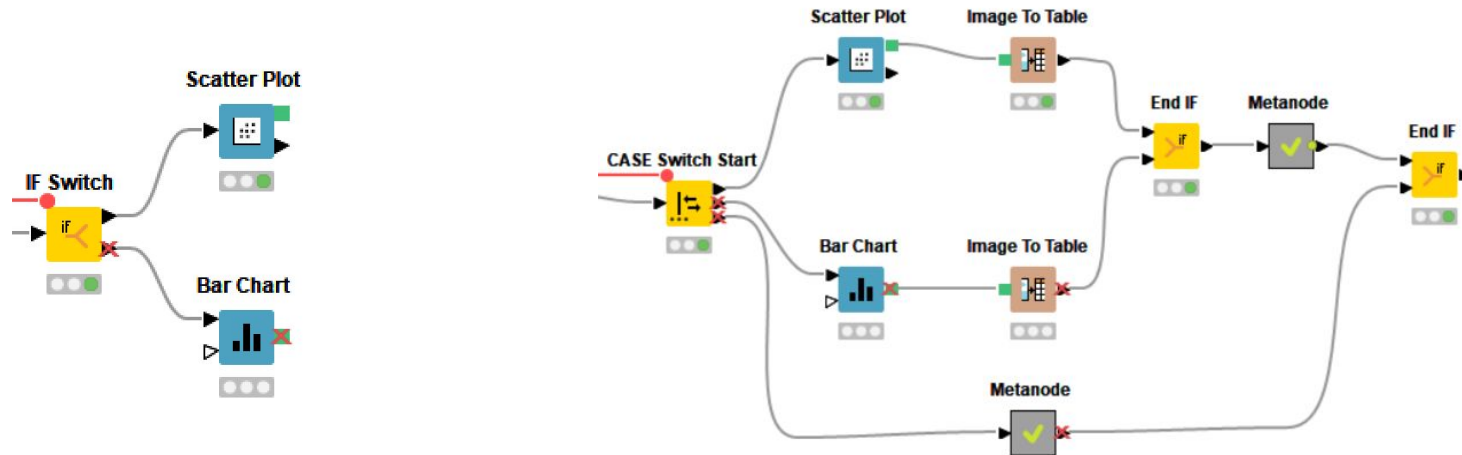
The Difference between Loops and Switches

Loops

- The Loop Start is connected to the Loop End node; they form a pair.
- A loop iterates over a workflow part.

Switches

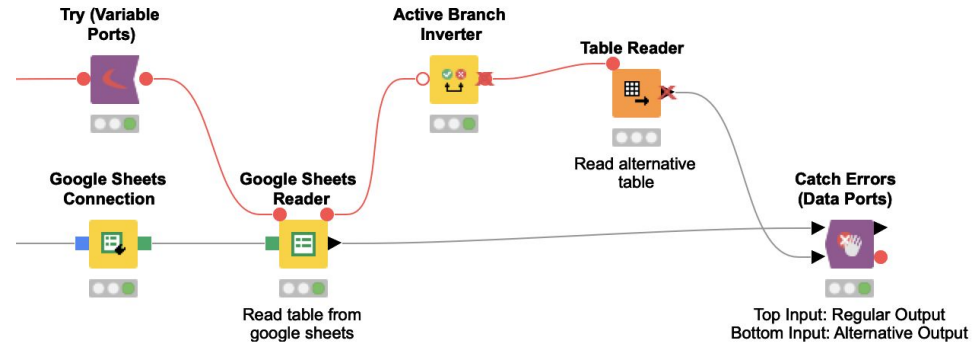
- A Switch Start can be used without a corresponding Switch End. They can also be combined.



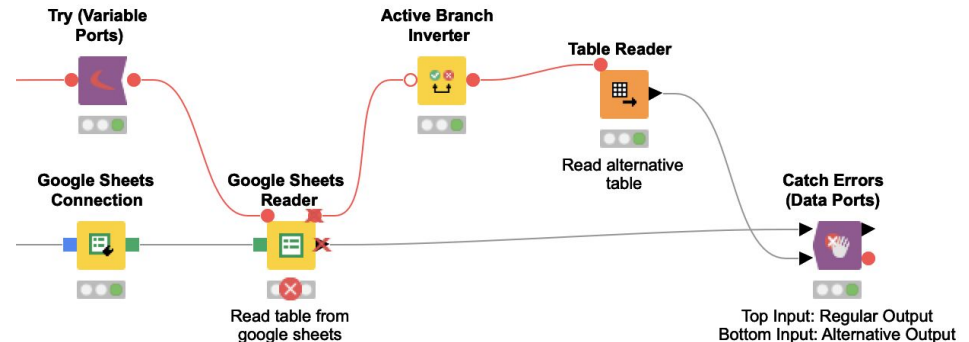
Try-Catch

- A way to catch errors in workflows
- Useful when it is hard to know if a node will execute (for example, when reading from a Google Sheet)
- KNIME tries to execute the nodes, but if it fails will fall back to an alternative branch

Regular Execution



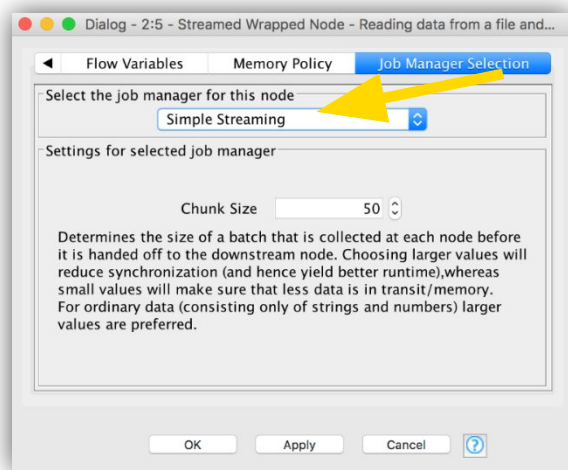
Alternative Execution



Streaming

- Standard execution: Node by node. The node processes all data, finishes, then passes the data to the next node, etc.
- Streaming: Nodes executed concurrently, each node passes data to the next as soon as it is available, i.e. before node is fully executed
 - Faster execution, esp. for reading/preprocessing data
- Install KNIME Streaming Execution (Beta) extension
- Create Component -> Configure -> Job Manager Selection -> Simple Streaming
 - Not available for all nodes (show in node repository)
 - Can only execute entire metanode, not individual nodes
 - Intermediate results not available since nothing is cached

Streaming



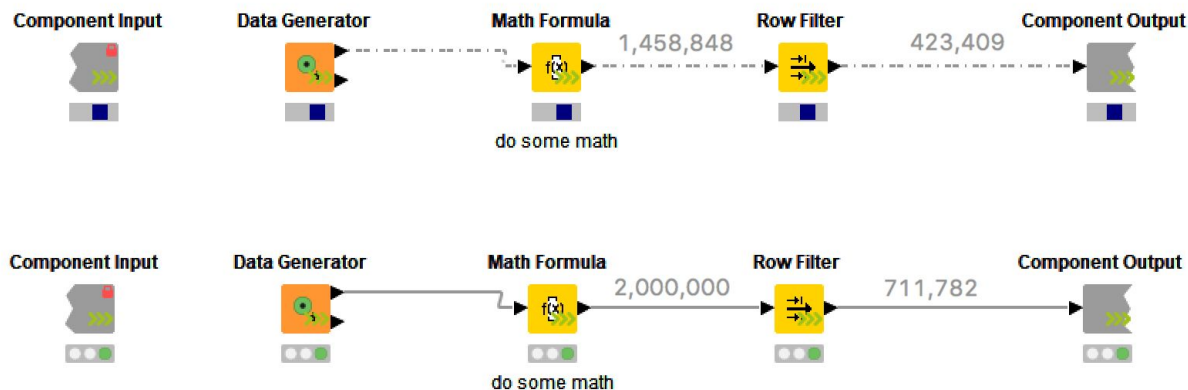
Streamed Component - Reading data from a file and process



Sub Workflow as Component.

To open it:

- right click > Component > Open
- Ctrl + Double Click



Exercise: 11_Workflow_Control, Activity I

Goal: Build a loop that will create an Excel file with separate Excel sheets for the records of different products.

- Read the table CurrentDetailData.table (Table Reader node)
- Start a loop that handles the records for the different products in separate iterations (Group Loop Start node)
- For each product write one Excel sheet into a single Excel file (Excel Writer node)
- Close and execute the loop (Variable Loop End node)

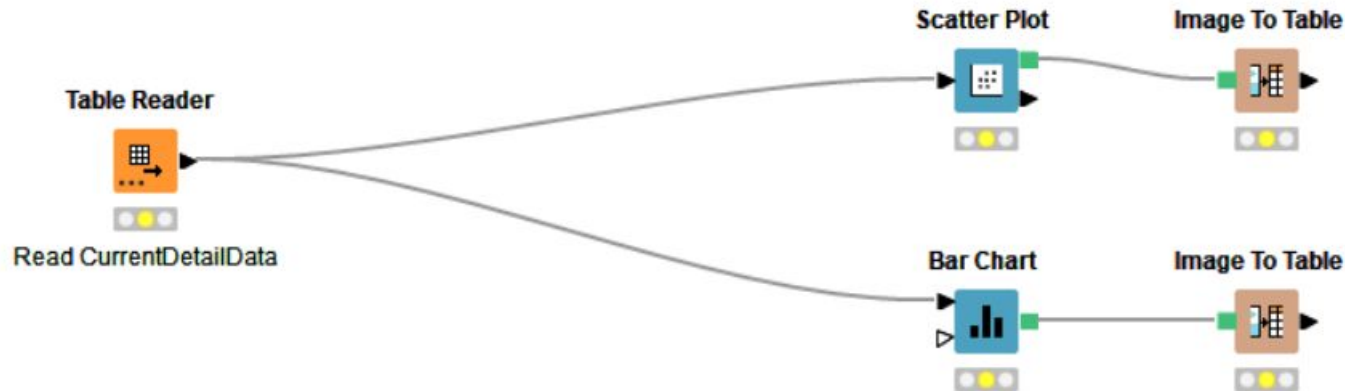
Exercise: 11_Workflow_Control, Activity II

Goal: Create a loop that reads and concatenates all the sheets in an Excel file.

- Create a table that contains all sheet names of the Excel file created in Activity I (Read Excel Sheet Names node)
- Start a loop that iterates over the sheet names (Table Row to Variable Loop Start node)
- Read the Excel sheet with the sheet name in the current iteration (Excel Reader node)
- Close the loop and concatenate the tables from the different iterations (Loop End node)

Exercise: 11_Workflow_Control, Activity III

- Extend the workflow below with a switch that only creates one type of visualization
 - Create a Single Selection Configuration node with the possible values "scatter" and "bar"
 - Use the CASE Switch Data (Start) that activates the top or the middle branch depending on the selection scatter/bar (Use the "... (index)" flow variable to define the active port)
 - Combine the outputs of the two branches with the CASE Switch Data (End) node



Instructor-Led Courses: Schedule 2023

Sep 11 - 16	L1	L1-DW KNIME Analytics Platform for Data Wranglers: Basics	Data Analysts
Sep 18 - 22	L2	L2-DW KNIME Analytics Platform for Data Wranglers: Advanced	
Oct 23 - 27	L3	L3-DA Productionizing Data Apps*	
Nov 6 - 10	L4	L4-DV Low Code Data Extraction and Visualization*	

Jun 19 - 23	L1	L1-DS KNIME Analytics Platform for Data Scientists: Basics	Data Scientists
Jun 26 - 30	L2	L2-DS KNIME Analytics Platform for Data Scientists: Advanced	
Jul 10 - 14	L3	L3-CD Continuous Deployment and MLOps*	
Jul 17 - 21	L4	L4-ML Introduction to Machine Learning Algorithms	
Jul 24 - 28	L4	L4-DL Introduction to Neural Networks and Deep Learning	
Aug 21 - 25	L4	L4-TS Introduction to Time Series Analysis*	
Aug 28 - Sep 1	L4	L4-TP Introduction to Text Processing	

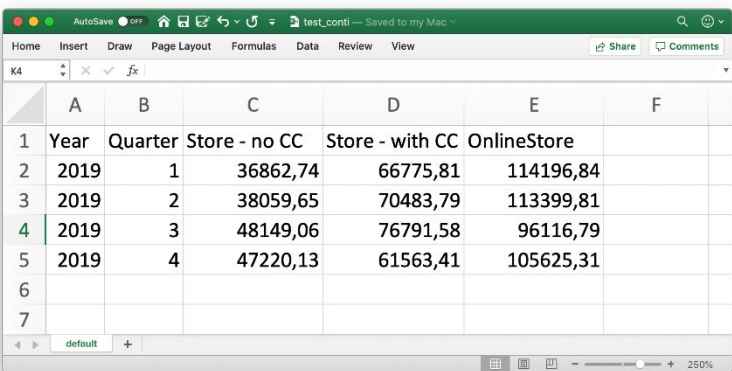
Sep 11 - 16	L1	L1-DW KNIME Analytics Platform for Data Wranglers: Basics	Data Engineers
Sep 18 - 22	L2	L2-DW KNIME Analytics Platform for Data Wranglers: Advanced	
Oct 9 - 13	L3	L3-DE Productionizing Data Engineering Applications*	
Oct 16 - 20	L4	L4-DE Best Practices for Data Engineering*	

Styling EXCEL Tables



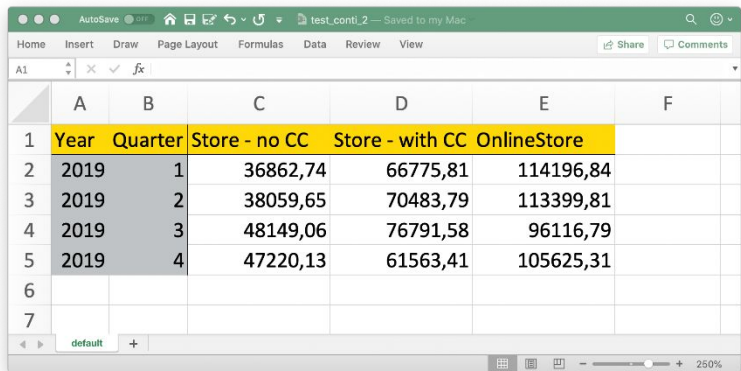
Styling EXCEL tables

Standard written table



	A	B	C	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						

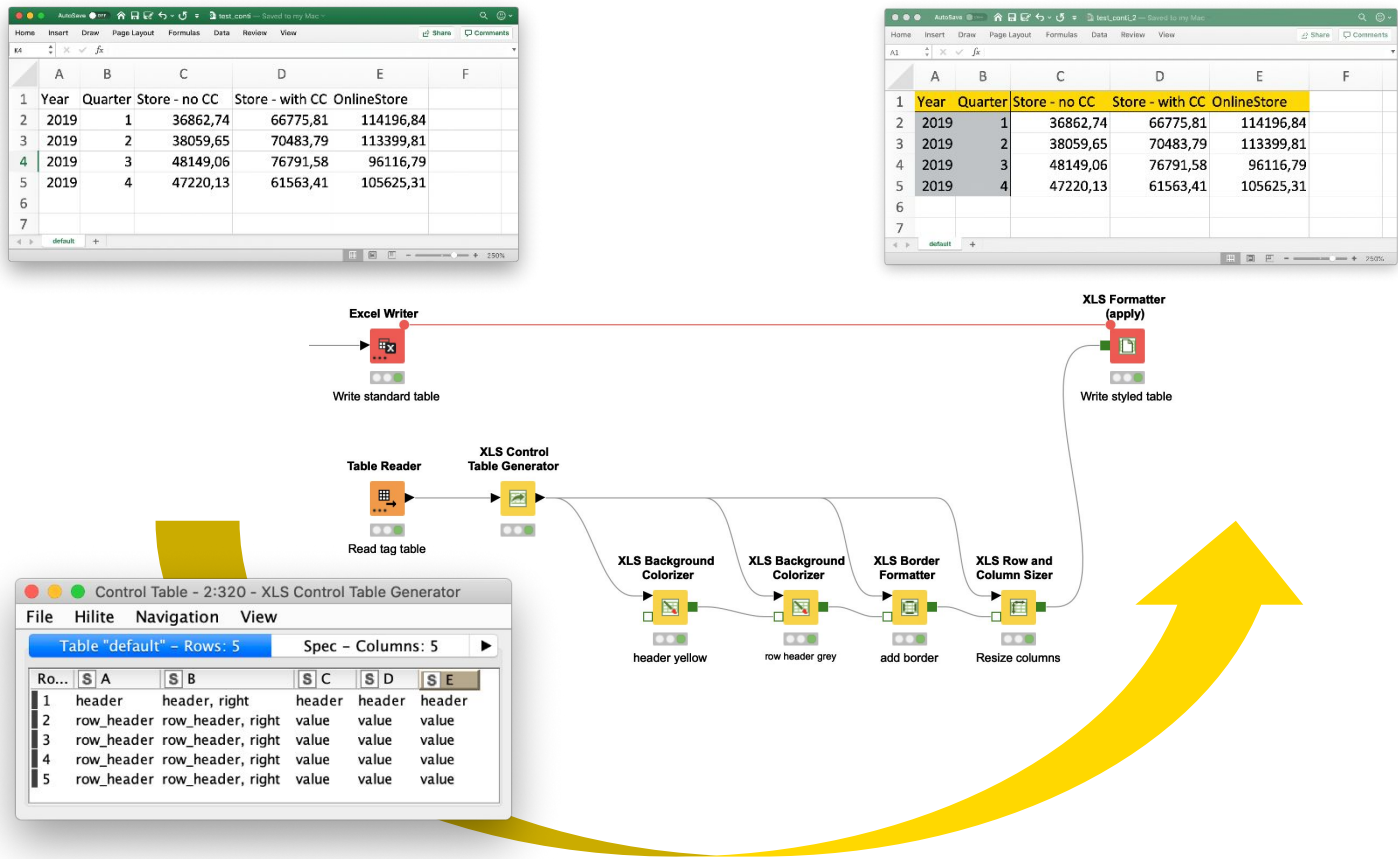
Styled table



	A	B	C	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						

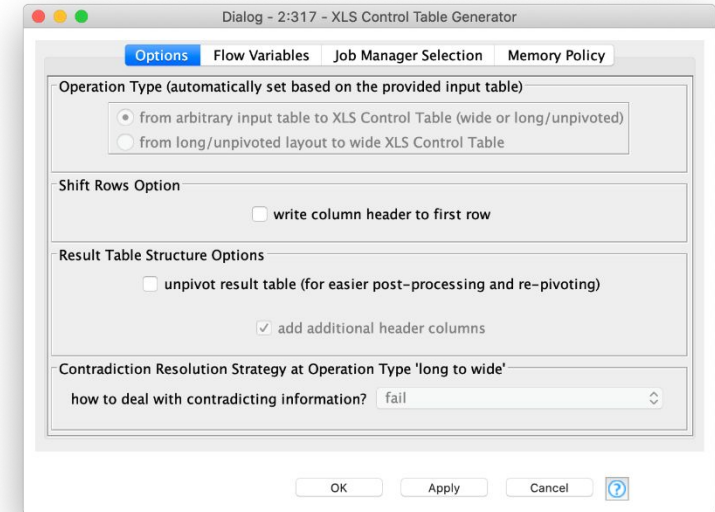
Continental Nodes for KNIME
XLS Formatter Nodes

Additional tag table - the key to a styled table:

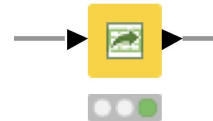


XLS Control Table Creator

- Kick off node to your styled table
- Transforms input table into an XLS Control Table
- Column names => A, B, C,
- Row IDs => 1, 2, 3, ...
- The unpivoted table is a great base to transform values into tags, e.g. with the Rule Engine node

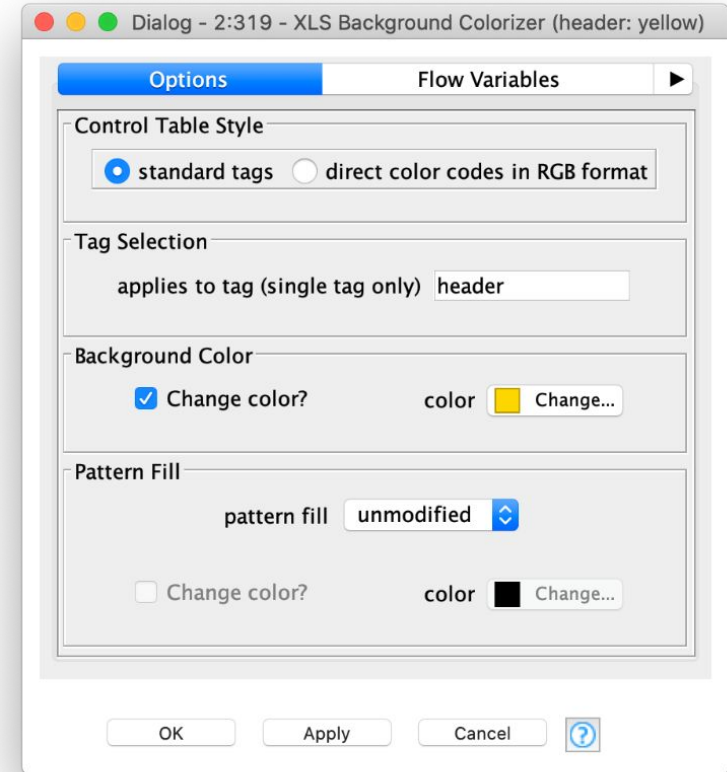
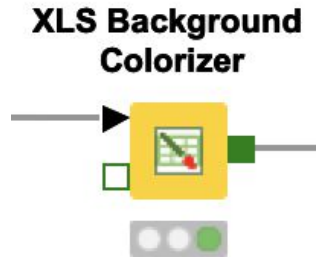


**XLS Control
Table Generator**



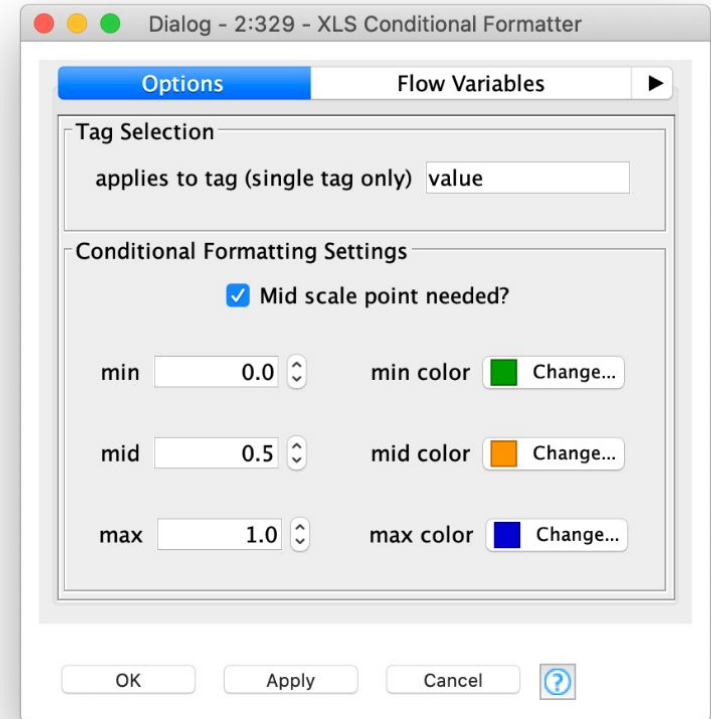
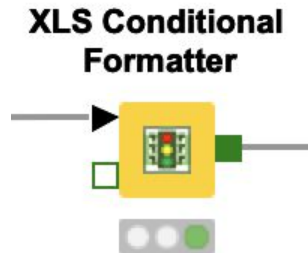
XLS Background Colorizer

- Changes the background of a cell into a static color and/or pattern fill
- Select between
 - Assigning a color to a tag value
 - Using a table with RGB values



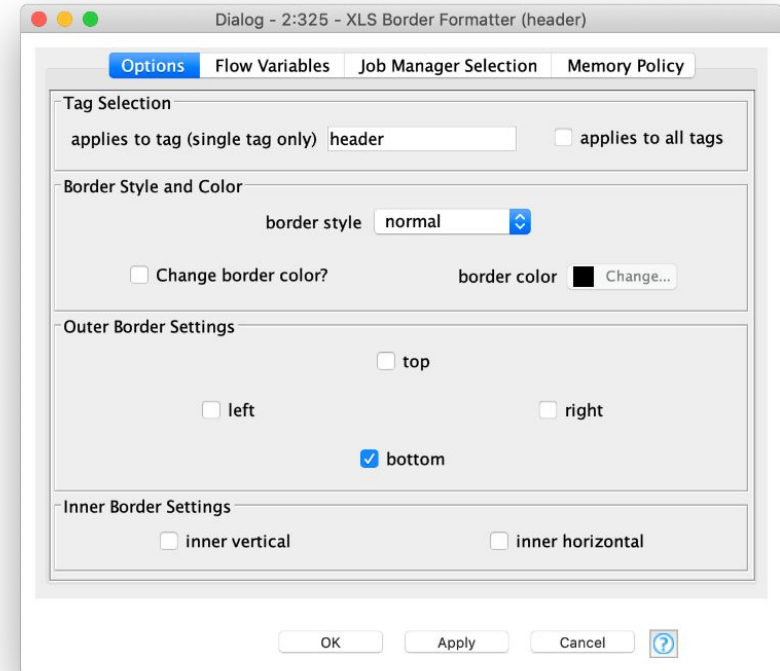
XLS Conditional Formatter

- Changes the cells' background according to their numerical value
- Allows you to optionally define a mid scale point



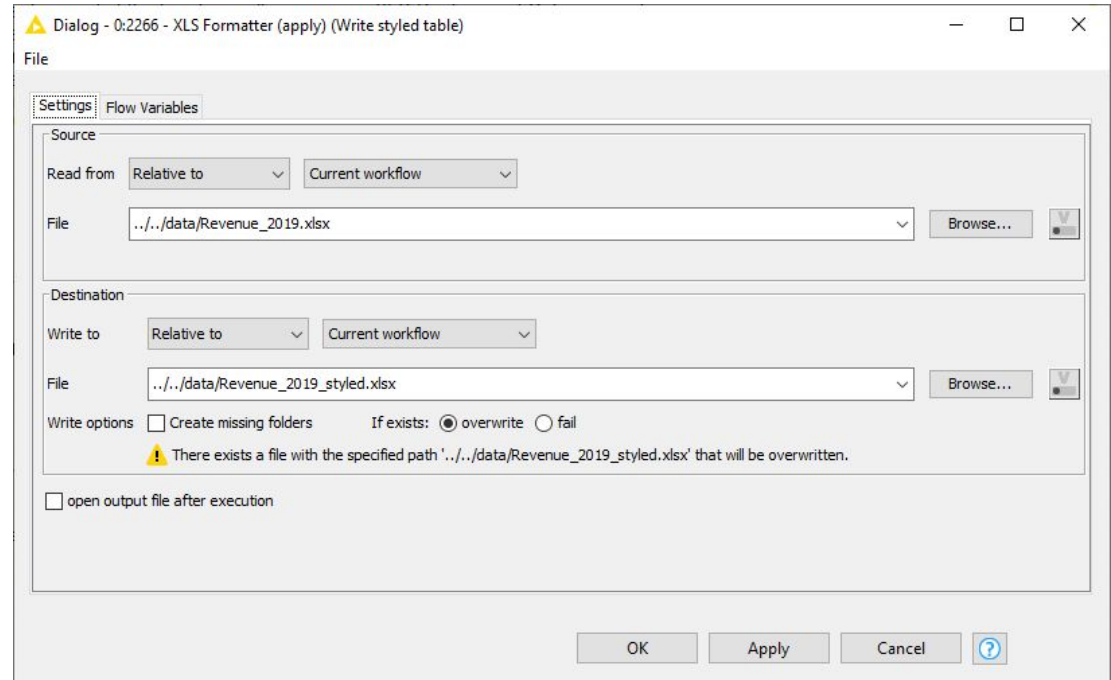
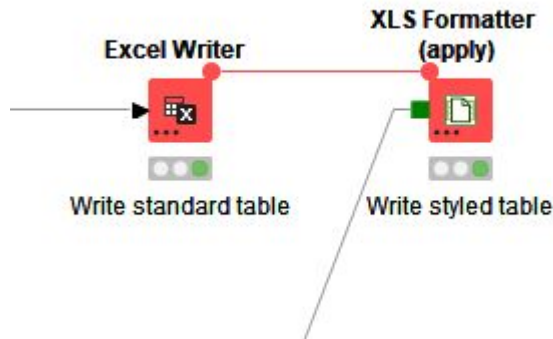
XLS Border Formatter

- Changes the borders of a given range specified by tags
- Option to use borders inside the specified range
 - Inner vertical
 - Inner horizontal

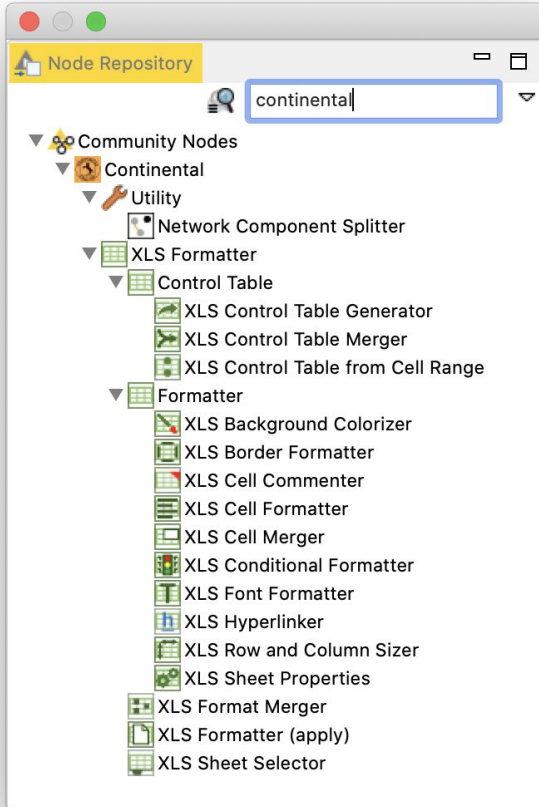


XLS Formatter (apply)

- Applies the chained commands from the XLS Formatter nodes
- Important:
 - The table must be written before with Excel Writer node
 - The input file may not contain any formatting yet

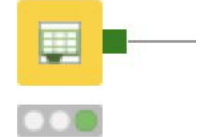


Other Important Nodes of the Extension



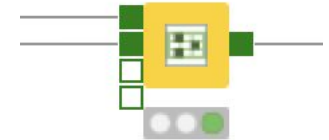
By default the styling is always applied to the first sheet. Change this with the XLS Sheet Selector node.

XLS Sheet Selector

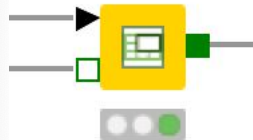


Want to style more than one sheet? Merge different sheet stylings with the XLS Format Merger node.

XLS Format Merger



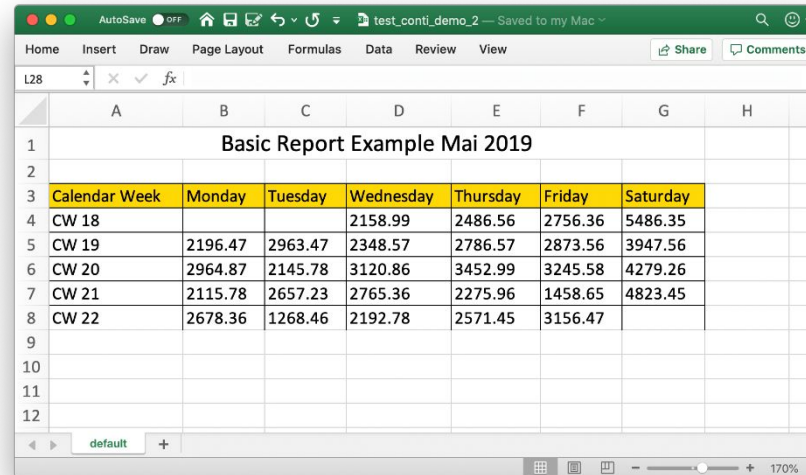
XLS Cell Merger



Want to add a header to your table? Merge multiple cells with the XLS Cell Merger node.

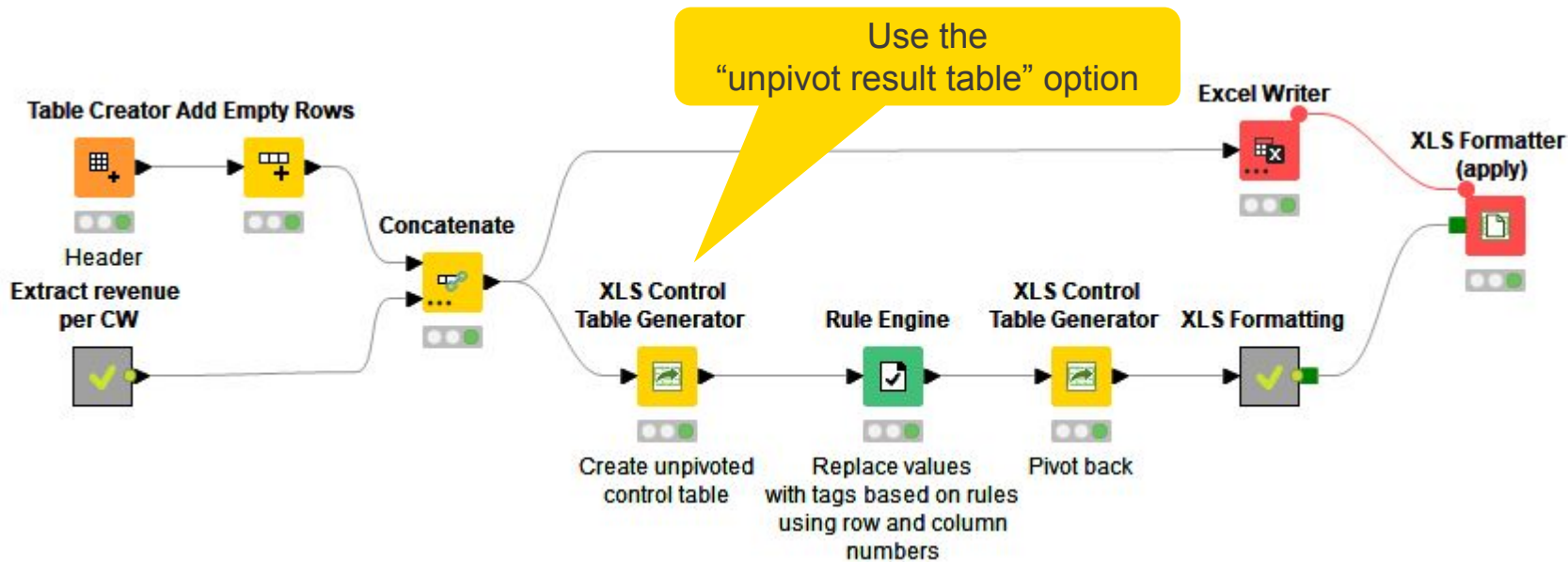
Tips & Tricks for the Continental Extension

- How can you automate creating a tag table?
 - Handle changing number of rows
 - Handle changing number of columns
- How can you add empty rows between different tables?



	A	B	C	D	E	F	G	H
1	Basic Report Example Mai 2019							
2								
3	Calendar Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	
4	CW 18			2158.99	2486.56	2756.36	5486.35	
5	CW 19	2196.47	2963.47	2348.57	2786.57	2873.56	3947.56	
6	CW 20	2964.87	2145.78	3120.86	3452.99	3245.58	4279.26	
7	CW 21	2115.78	2657.23	2765.36	2275.96	1458.65	4823.45	
8	CW 22	2678.36	1268.46	2192.78	2571.45	3156.47		
9								
10								
11								
12								

Tips & Tricks for the Continental extension



Tips & Tricks for the Continental extension

Table Creator Add



Header

Extract
pe

Control Table - 5:2260 - XLS Control Table Generator (Create unpivoted)

File Hilite Navigation View

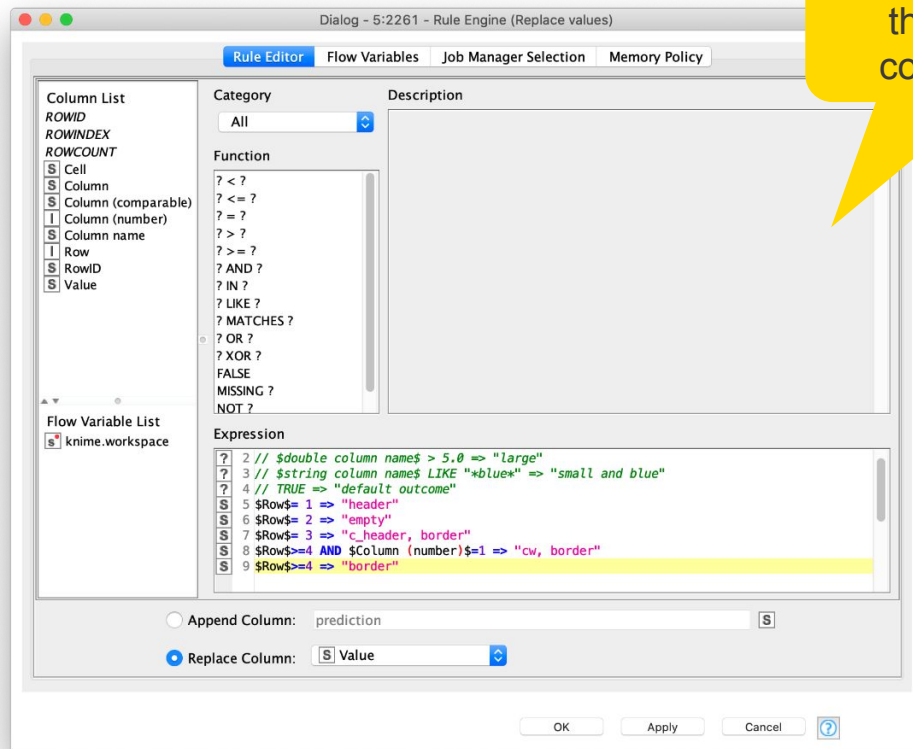
Table "default" - Rows: 56 Spec - Columns: 8 Properties Flow Variables

Row ID	S Cell	S Column	S Column (comparable)	I Column (number)	S Column name	I Row	S RowID	S Value
Row0	A1	A	00A	1	column1	1	Row0	Basic Report Example Mai 2019
Row1	B1	B	00B	2	column2	1	Row0	?
Row2	C1	C	00C	3	column3	1	Row0	?
Row3	D1	D	00D	4	column4	1	Row0	?
Row4	E1	E	00E	5	column5	1	Row0	?
Row5	F1	F	00F	6	column6	1	Row0	?
Row6	G1	G	00G	7	column7	1	Row0	?
Row7	A2	A	00A	1	column1	2	Empty 0	?
Row8	B2	B	00B	2	column2	2	Empty 0	?
Row9	C2	C	00C	3	column3	2	Empty 0	?
Row10	D2	D	00D	4	column4	2	Empty 0	?
Row11	E2	E	00E	5	column5	2	Empty 0	?
Row12	F2	F	00F	6	column6	2	Empty 0	?
Row13	G2	G	00G	7	column7	2	Empty 0	?
Row14	A3	A	00A	1	column1	3	Row0_dup	Calendar Wee
Row15	B3	B	00B	2	column2	3	Row0_dup	Monday
Row16	C3	C	00C	3	column3	3	Row0_dup	Tuesday
Row17	D3	D	00D	4	column4	3	Row0_dup	Wednesday
Row18	E3	E	00E	5	column5	3	Row0_dup	Thursday
Row19	F3	F	00F	6	column6	3	Row0_dup	Friday
Row20	G3	G	00G	7	column7	3	Row0_dup	Saturday
Row21	A4	A	00A	1	column1	4	Row1	CW 18
Row22	B4	B	00B	2	column2	4	Row1	?
Row23	C4	C	00C	3	column3	4	Row1	?
Row24	D4	D	00D	4	column4	4	Row1	2158.99
Row25	E4	E	00E	5	column5	4	Row1	2486.56
Row26	F4	F	00F	6	column6	4	Row1	2756.36
Row27	G4	G	00G	7	column7	4	Row1	5486.35

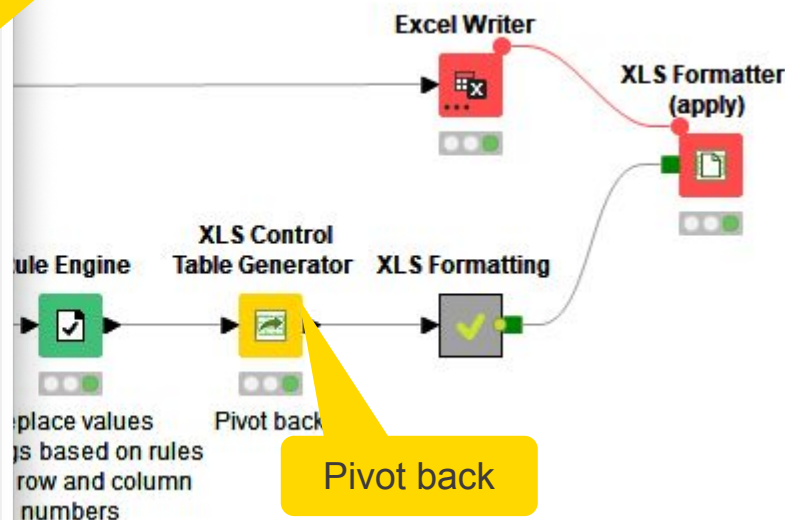
formatter
ply)

For each cell one row
with row and column
number, and Cell ID

Tips & Tricks for the Continental extension



Define rules based on the row and the column numbers

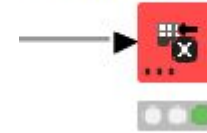


Pivot back

Excel Cell Updater

- Updates values of cells at specified address in an Excel file
 - While retaining the cell format

Excel Cell Updater



- Provide an input table of cell addresses and values to be replaced

- Example

Address	String	Integer
A5	Ok	?
5:96	?	50
OZ23914	?	?

- Missing values are denoted by “?”
- Cell A5 is replaced with a string “OK”
- Cell 5-th row, 96-th column is replaced with an integer 50
- Cell OZ23914 is cleared

Excel Cell Updater

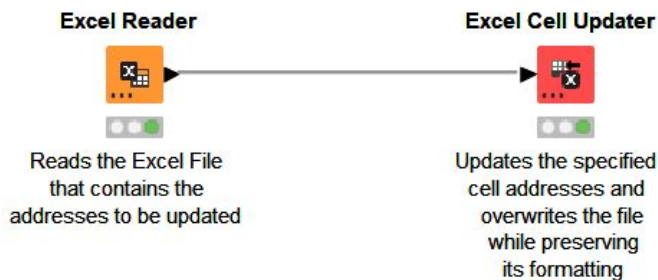


Table with replacement values

Row ID	S Address	S String	S Integer
Row0	H13	yes	?
Row1	H16	yes	?

	A	B	C	D	E	F	G	H	
1	age	job	marital	education	default	balance	housing	loan	cont.
2	58	managem	married	tertiary	no	2143	yes	no	unkr
3	44	techniciar	single	secondary	no	29	yes	no	unkr
4	33	entrepren	married	secondary	no	2	yes	yes	unkr
5	47	blue-colla	married	unknown	no	1506	yes	no	unkr
6	33	unknown	single	unknown	no	1	no	no	unkr
7	35	managem	married	tertiary	no	231	yes	no	unkr
8	28	managem	single	tertiary	no	447	yes	yes	unkr
9	42	entrepren	divorced	tertiary	yes	2	yes	no	unkr
10	58	retired	married	primary	no	121	yes	no	unkr
11	43	techniciar	single	secondary	no	593	yes	no	unkr
12	41	admin.	divorced	secondary	no	270	yes	no	unkr
13	29	admin.	single	secondary	no	390	no	no	unkr
14	53	techniciar	married	secondary	no	6	yes	no	unkr
15	58	techniciar	married	unknown	no	71	yes	no	unkr
16	57	services	married	secondary	no	162	no	no	unkr

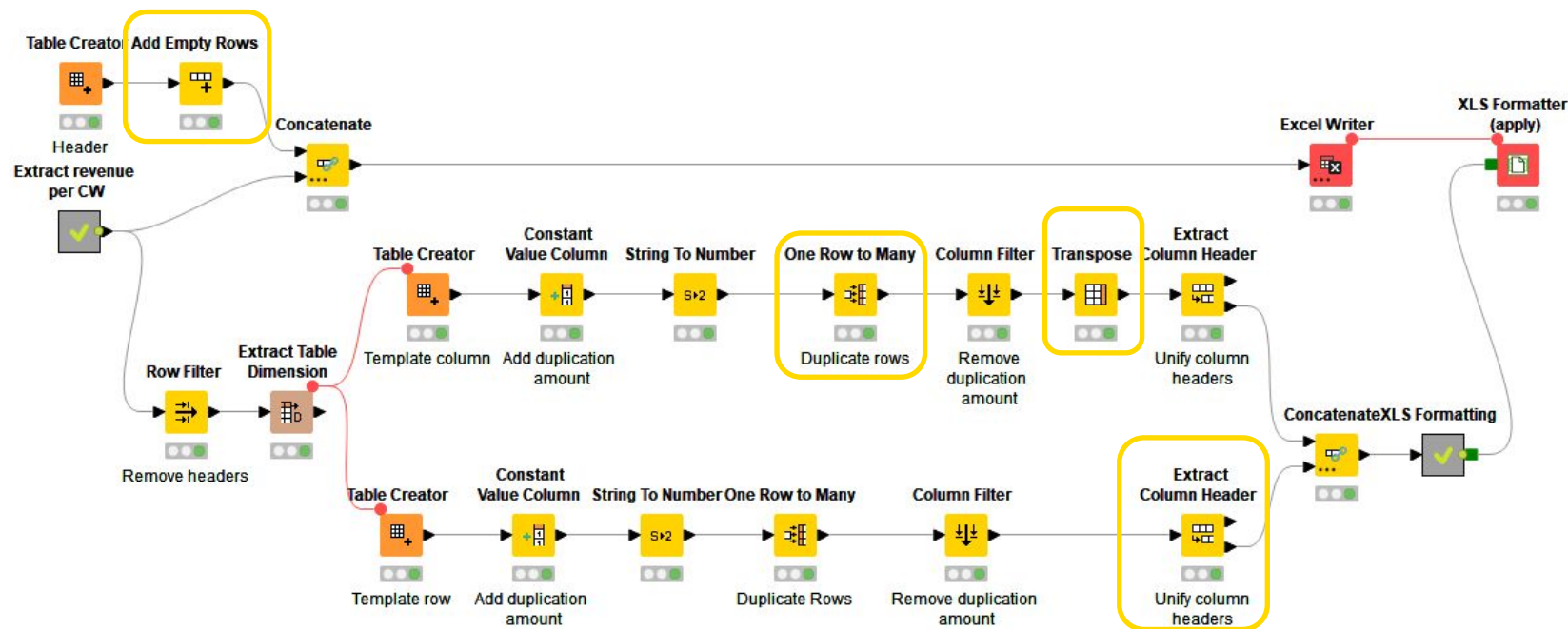
Updated
Original Excel file



	G	H	
using	loan	cont	
s	no	unkr	
s	no	unkr	
s	yes	unkr	
s	no	unkr	
i	no	unkr	
s	no	unkr	
s	yes	unkr	
s	no	unkr	
s	no	unkr	
s	no	unkr	
s	no	unkr	
i	yes	unkr	
s	no	unkr	
s	no	unkr	
i	yes	unkr	

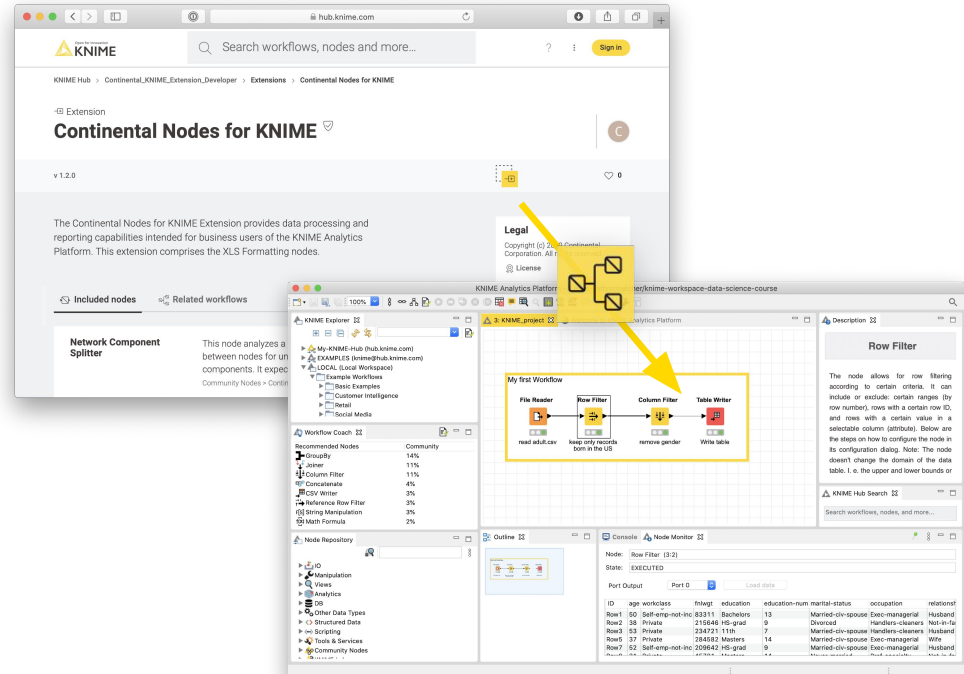
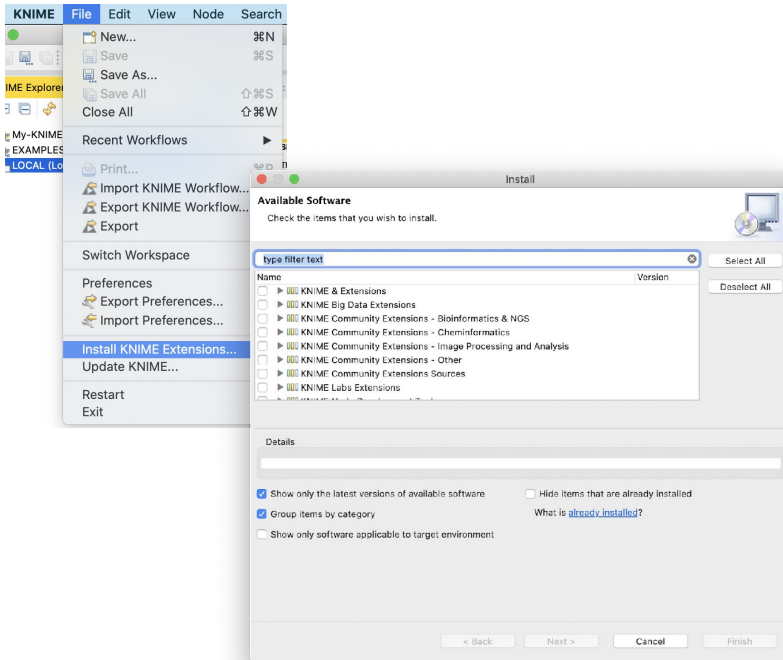
Conditional formatting
is preserved

Tips & Tricks for the Continental extension



Installing Extensions

- Install the Continental Nodes for KNIME extension
- Install extension by going to File -> Install KNIME Extension or via Drag & Drop from the KNIME Community Hub

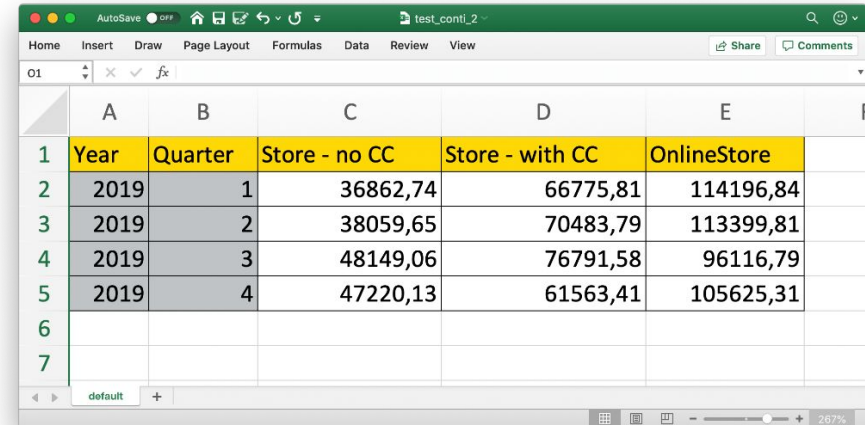


Exercise: 12_Styling_Excel_Tables

- Read the files: TagTable.table and RevenuePerQuarter.table and have a look at the tables.
- Use the XLS Formatter node to change the styling of the RevenuePerQuarter.table table to look like this, below:

Hint: Use the following nodes:

- XLS Control Table Generator
- XLS Background Colorizer
- XLS Border Formatter
- XLS Row and Columns Sizer (optional)



	A	B	C	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						

Introduction to Data Science



Churn Prediction



CRM System
Data about your customer

- Demographics
- Behavior
- Revenues



Model



- Churn Prediction
- Upselling Likelihood
- Product Propensity /NBO
- Campaign Management
- Customer Segmentation
- ...

Customer Segmentation



CRM System
Data about your customer

- Demographics
- Behavior
- Revenues



Model

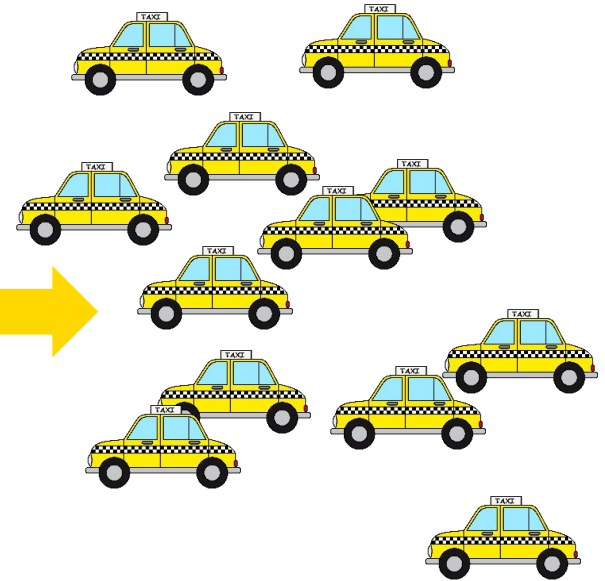


- Churn Prediction
- Upselling Likelihood
- Product Propensity /NBO
- Campaign Management
- **Customer Segmentation**
- ...

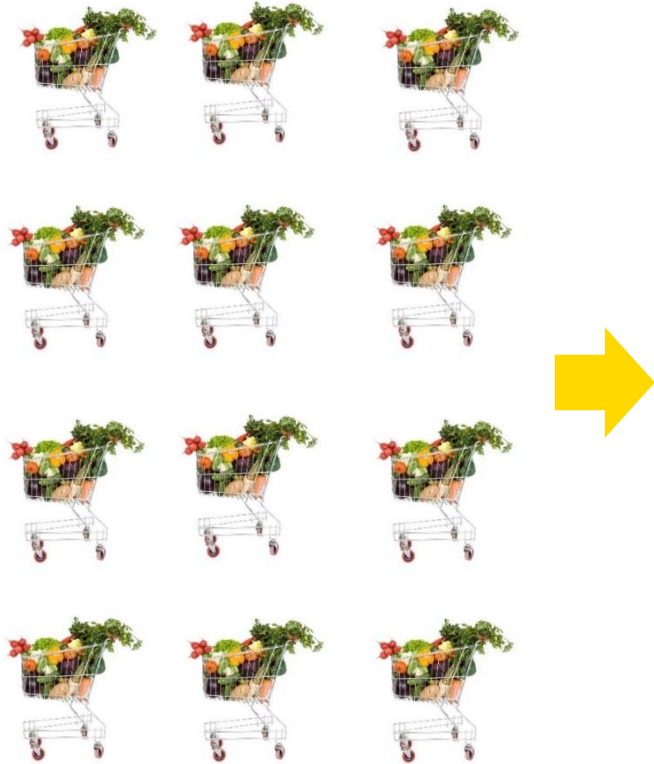
Demand Prediction



How many taxis
do I need in NYC
on Wednesday at
12:00?



Recommendation Engines / Market Basket Analysis



Model

Rule: $X \Rightarrow Y$

$$\text{Support} = \frac{\text{freq}(X, Y)}{N}$$
$$\text{Confidence} = \frac{\text{freq}(X, Y)}{\text{freq}(X)}$$
$$\text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)}$$



Recommendation

IF



=>



Sentiment Analysis



Samsung

Samsung Galaxy S7 Edge G935A 32GB Unlocked - Gold Platinum



125 customer reviews | 606 answered questions



Beautiful phone from a wonderful seller!

By

on May 29, 2017

Color: Gold | **Verified Purchase**

This practically new beautiful phone well exceeded my expectations!



One Star

By

on August 3, 2016

Color: Black Onyx | **Verified Purchase**

Very bad experience



Today's Challenge – Analyze Some Customer Data

Initial situation:

- Customer data in two datasets:
 - Phone usage
 - Contract information
- Column “Phone” is in both datasets
- Column “churn” encodes whether a customer is happy

Goal:

- Find rules that describe happy and unhappy customers by
 - training a decision tree model
 - calculating aggregations (optional)

Classification

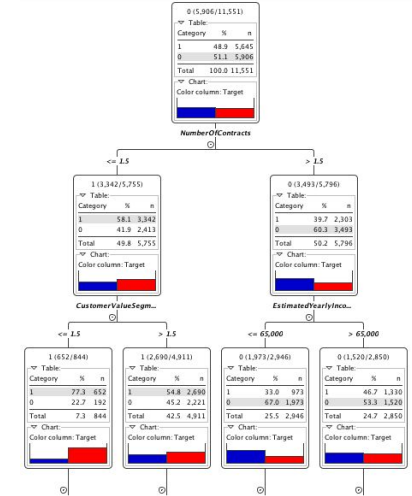
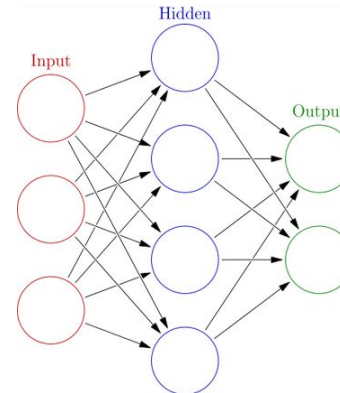
Predict *nominal* outcomes on existing data (supervised)

■ Applications

- Churn analysis (yes/no)
- Chemical activity (active/inactive)
- Spam detection (spam/not spam)
- Optical character recognition (A-Z)

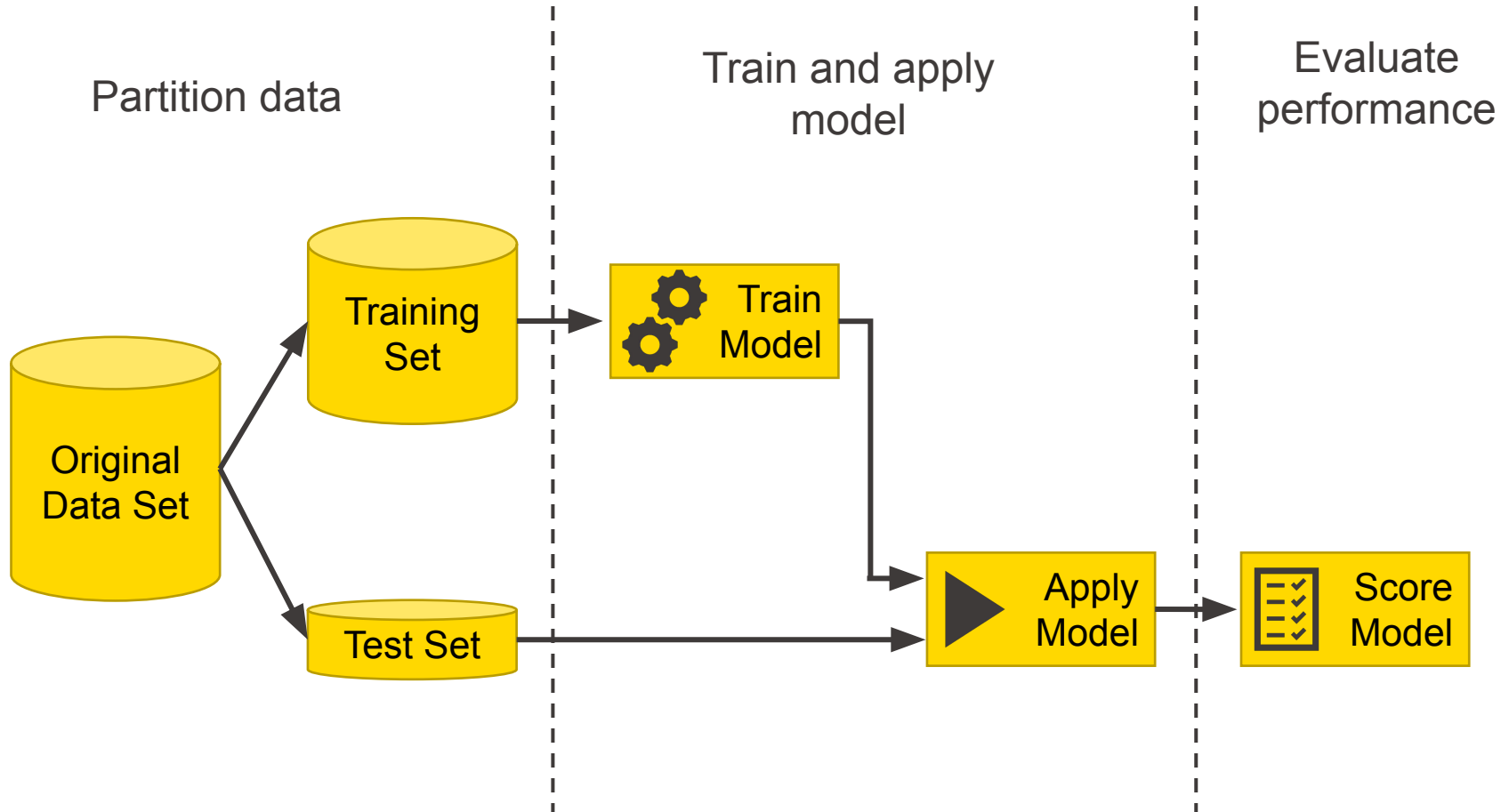
■ Methods

- Decision Trees
- Neural Networks
- Naïve Bayes
- Logistic Regression



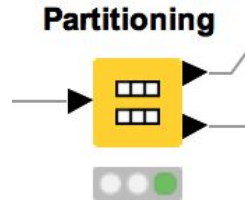
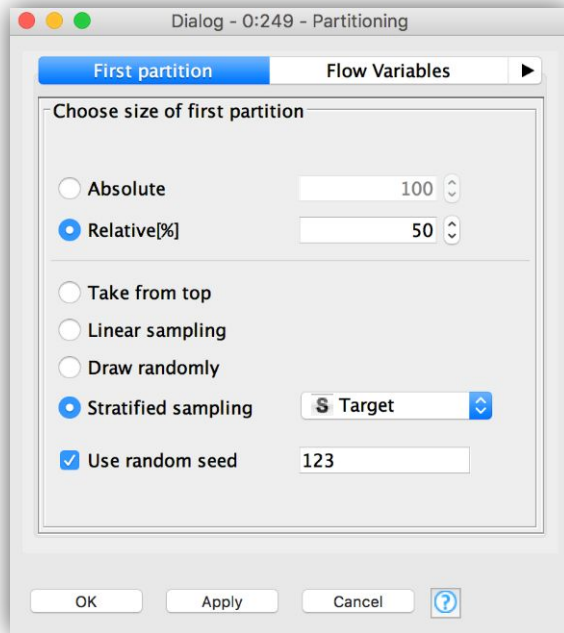
Class counts for Target		
Class:	0	1
Count:	5906	5645
Total count: 11551		
Threshold to used for zero probabilities: 0.0		
Gaussian distribution for Age per class value		
	0	1
Count:	5906	5645
Mean:	49.68557	46.82604
Std. Deviation:	12.27388	10.16363
Rate:	51%	49%
Gaussian distribution for Available401K per class value		
	0	1
Count:	5906	5645
Mean:	0.68134	0.68485
Std. Deviation:	0.466	0.46462
Rate:	51%	49%

Data Mining: Process Overview



Partitioning

- Use to split data into training and evaluation sets
 - Partition by count (e.g. 10 rows) or fraction (e.g. 10%)
 - Sample by a variety of methods; random, linear, stratified



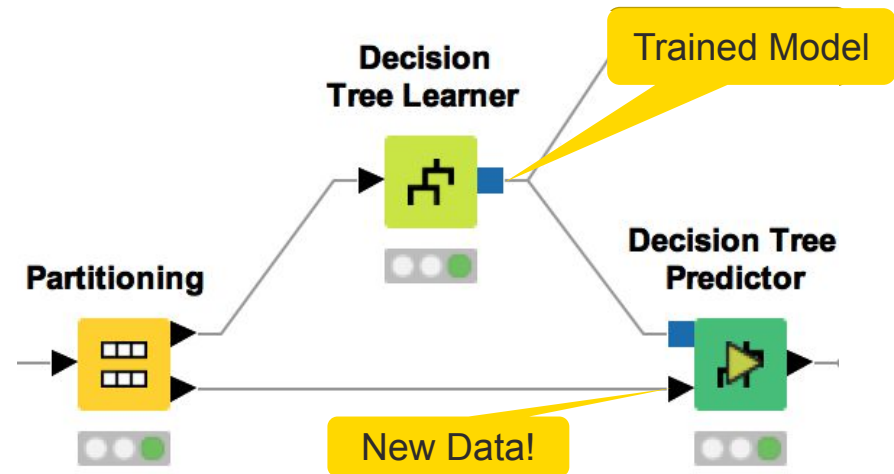
The first screenshot shows the 'First partition (as defined in dialog) - 0:249 - Partitioning' table with 5775 rows. The second screenshot shows the 'Second partition (remaining rows) - 0:249 - Partitioning' table with 5776 rows. Both tables have columns for Row ID, Marital Status, Gender, Estimated Value, Number of Children, and Age.

Row ID	Marital Status	Gender	Estimated Value	Number of Children	Age
Row0	M	M	90000	0	44
Row7	M	M	60000	2	46
Row9	S	M	70000	1	46
Row10	S	F	70000	1	46
Row13	M	M	100000	3	42
Row14	S	F	100000	3	42
Row15	S	F	30000	1	31
Row17	S	F	20000	2	66
Row18	S	M	30000	2	66
Row20	S	M	40000	2	32

Row ID	Marital Status	Gender	Estimated Value	Number of Children	Age
Row1	S	M	60000	1	45
Row2	M	M	60000	1	45
Row3	S	F	70000	1	42
Row4	S	F	80000	4	42
Row5	S	M	70000	1	45
Row6	S	F	70000	1	44
Row8	S	F	60000	3	46
Row11	M	M	60000	4	46
Row12	M	F	100000	2	42
Row16	M	M	30000	1	31
Row19	S	M	40000	2	32

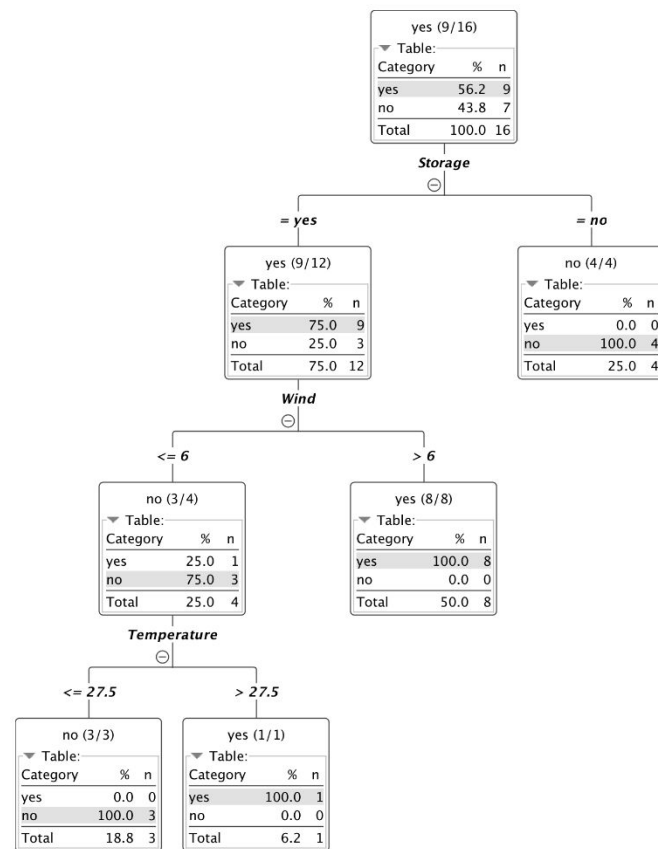
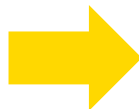
Learner-Predictor Motif

- Most data mining approaches in KNIME use a Learner-predictor motif.
- The Learner node trains the model with its input data.
- The Predictor node applies the model to a different subset of data.



Goal: A Decision Tree

Outlook	Wind	Temp	Storage	Sailing
sunny	3	30	yes	yes
sunny	3	25	yes	no
rain	12	15	yes	yes
overcast	15	2	no	no
rain	16	25	yes	yes
sunny	14	18	yes	yes
rain	3	5	no	no
sunny	9	20	yes	yes
overcast	14	5	no	no
sunny	1	7	no	no
rain	4	25	yes	no
rain	14	24	yes	yes
sunny	11	20	yes	yes
sunny	2	18	yes	no
overcast	8	22	yes	yes
overcast	13	24	yes	yes



Decision Tree Learner



Dialog - 0:277 - Decision Tree Learner

Options PMMLSettings Flow Variables ▶

General

Class column

Quality measure

Pruning method

☒ Reduced Error Pruning

Min number records per node

Number records to store for view

☒ Average split point

Number threads

☒ Skip nominal columns without domain information

Root split

☐ Force root split column

Root split column

Binary nominal splits

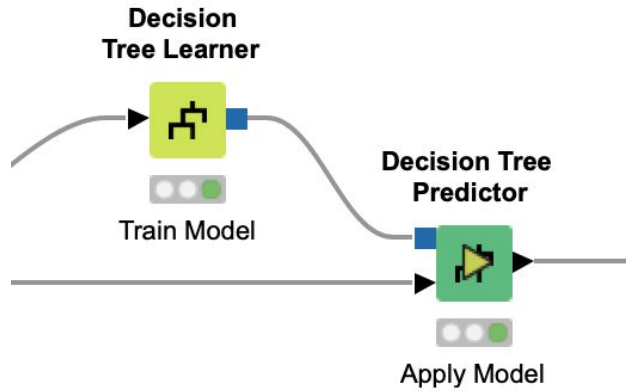
☐ Binary nominal splits

Max #nominal

☐ Filter invalid attribute values in child nodes

OK Apply Cancel ?

Applying the Model – What are the Outputs?



Classified Data - 0:65 - Decision Tree Predictor (Apply Model)

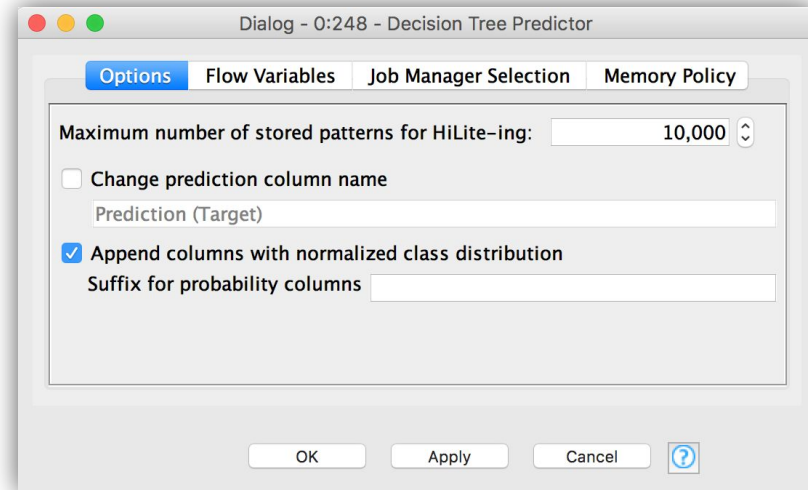
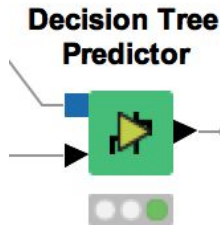
File Hilite Navigation View

Table "default" - Rows: 879 Spec - Columns: 82 Properties Flow Variables

Row ID	..	I	SalePr...	S rank	D P (rank=Low)	D P (rank=High)	S Prediction (rank)
10			189000	Low	0.889	0.111	Low
11			175900	Low	1	0	Low
13			180400	Low	1	0	Low
15			212000	Low	0.946	0.054	Low
21			190000	High	0	1	High
22			170000	High	0.2	0.8	High
27			126000	Low	1	0	Low
28			115000	Low	1	0	Low
33			127500	Low	1	0	Low

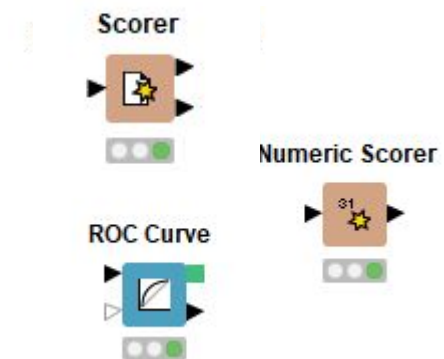
Decision Tree Predictor

- Takes a decision tree model and apply it to new data
- Check the box to append class probabilities



Evaluation Metrics

- Why evaluation metrics?
 - Quantify the power of the model as a classifier/predictor
 - Compare model configurations and/or models, and select the best performing one
 - Obtain the expected performance of the model for new data
- Different model evaluation techniques are available for
 - Classification/regression models
 - Imbalanced/balanced target class distributions



Overall Accuracy

$$\textit{Overall Accuracy} = \frac{\# \textit{Correct Classifications}}{\# \textit{All Events}}$$

- The proportion of correct classifications
- Downsides:
 - Only considers the performance in general and not for the different target classes
 - Therefore, not informative when the target class distribution is unbalanced

Confusion Matrix

Arbitrarily define one target class as POSITIVE and the remaining class(es) as NEGATIVE

	Predicted class POSITIVE	Predicted class NEGATIVE
Actual class POSITIVE	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
Actual class NEGATIVE	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)

TRUE POSITIVE (**TP**): Actual and predicted class is positive

TRUE NEGATIVE (**TN**): Actual and predicted class is negative

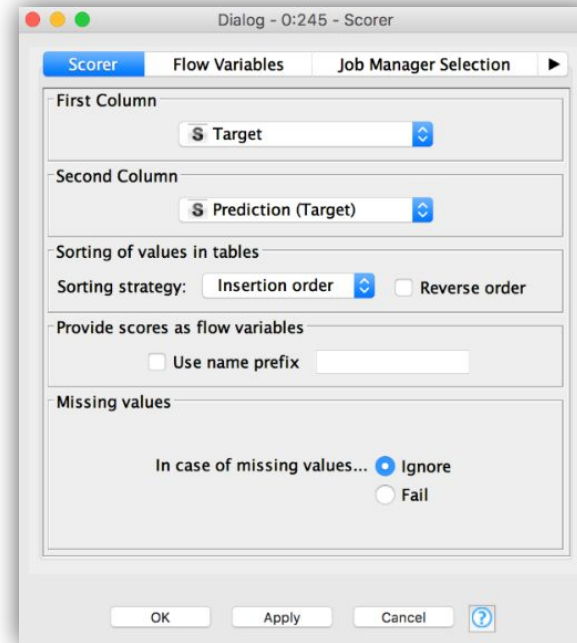
FALSE NEGATIVE (**FN**): Actual class is positive and predicted negative

FALSE POSITIVE (**FP**): Actual class is negative and predicted positive

- Use these four statistics to calculate other evaluation metrics, such as *overall accuracy*, *true positive rate*, and *false positive rate*

Scorer

Compare predicted results to known truth in order to evaluate model quality



Scorer

Confusion matrix shows the distribution of model errors

Confusion Matrix - 0:297 - Scorer

File

Hilite

Target \ Prediction (Target)	1	0
1	2073	750
0	759	2193

Correct classified: 4,266

Wrong classified: 1,509

Accuracy: 73.87 %

Error: 26.13 %

Cohen's kappa (k) 0.477

An accuracy statistics table provides a detailed analysis of model quality

Accuracy statistics - 0:297 - Scorer

File Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 11 Properties Flow Variables

Row ID	TruePositives	FalsePositives	TrueNegatives	FalseNegatives	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy	Cohen's kappa
1	2073	759	2193	750	0.734	0.732	0.734	0.743	0.733	?	?
0	2193	750	2073	759	0.743	0.745	0.743	0.734	0.744	?	?
Overall	?	?	?	?	?	?	?	?	?	0.739	0.477

Exercise: 13_Training_a_Churn_Prediction_Model

- Read the CallsData.xls and ContractData.csv files
- Join the two data tables based on the columns “Area Code” and “Phone”
- Change the data type of the columns “Area Code” and “Churn” to string
- Partition the data into a training set and a test set
- Train a decision tree to detect customers that are likely to churn
- Apply the mode to the test set and evaluate the model performance

Confirmation of Attendance and Survey

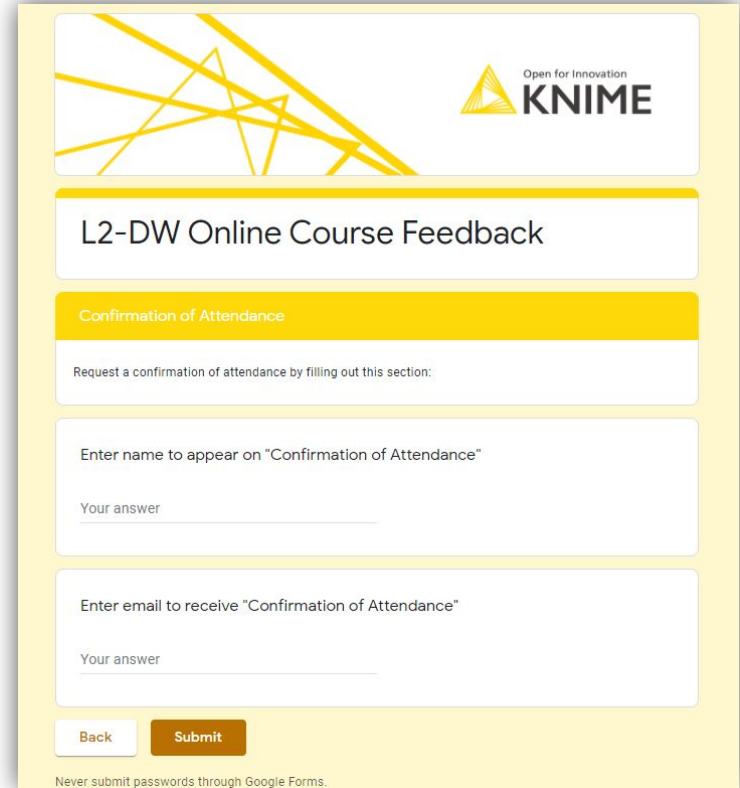
- If you would like to get a “Confirmation of Attendance” please click on the link below*

[Confirmation of Attendance and Survey](#)

- The link also takes you to our course feedback survey. Filling it in is optional but highly appreciated!

Thank you!

*Please send your request within the next 3 days



The screenshot shows a Google Forms interface for the 'L2-DW Online Course Feedback' survey. At the top, there is a header with the KNIME logo and the tagline 'Open for Innovation'. Below the header, the title 'L2-DW Online Course Feedback' is displayed. The main section is titled 'Confirmation of Attendance' and contains a prompt: 'Request a confirmation of attendance by filling out this section:'. There are two input fields: 'Enter name to appear on "Confirmation of Attendance"' and 'Enter email to receive "Confirmation of Attendance"', both with 'Your answer' placeholder text. At the bottom, there are 'Back' and 'Submit' buttons. A footer note states 'Never submit passwords through Google Forms.'

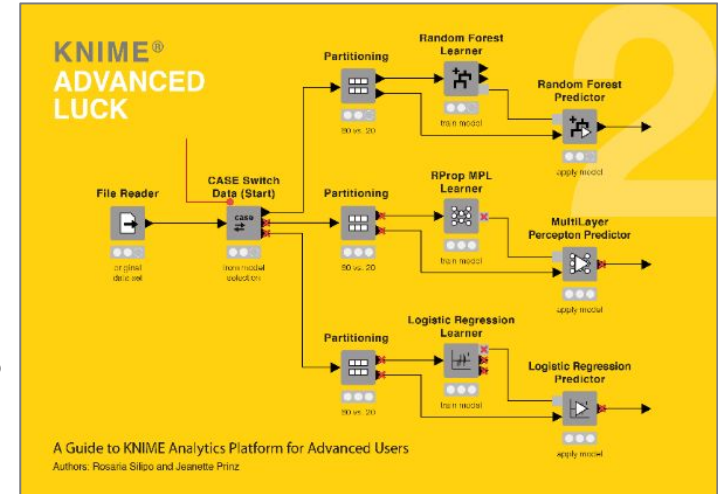
KNIME Books

Course books are available on
KNIME Press

<https://www.knime.com/knimepress>

Download your free copy of the KNIME Beginner's
Luck by using the code: **SUMMIT-BOOKS-0423**

<https://www.knime.com/knimepress/knime-advanced-luck>

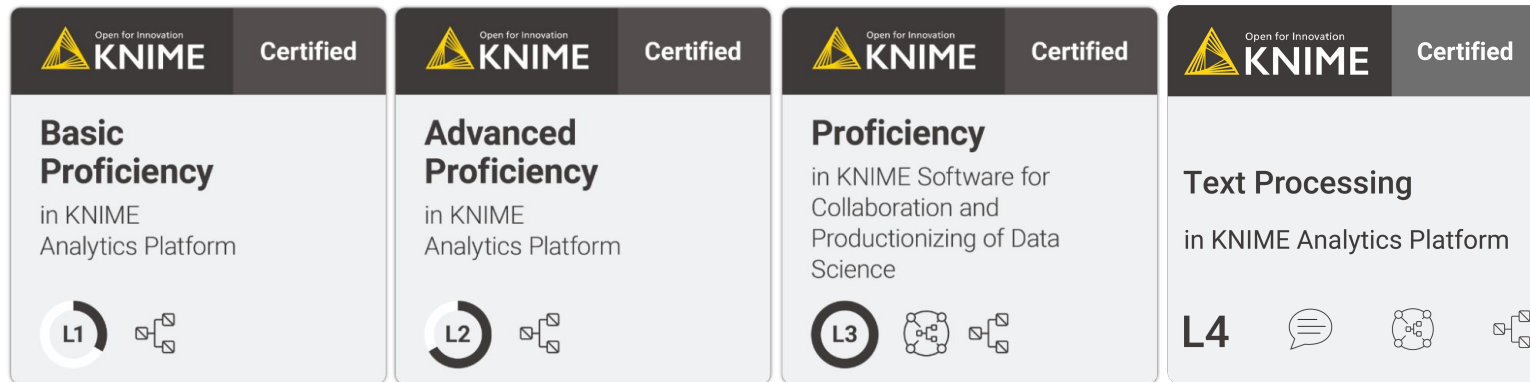


KNIME Certification Program

Measure and certify your KNIME and data science expertise

<https://www.knime.com/certification-program>

Four levels L1 – L4 of on-demand certification examinations



Free certification exams for a week (until April 24, 2023) with the codes:

L2= SUMMIT-CERT-L2-0423 L3= SUMMIT-CERT-L3-0423

L4= SUMMIT-CERT-L4-0423 (L1 is free)

Stay Connected with KNIME



Blog:
knime.com/blog



KNIME Self-Paced Courses:
knime.com/knime-self-paced-courses



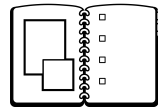
Forum:
forum.knime.com



Email:
education@knime.com



KNIME Hub:
hub.knime.com



Medium Journal:
medium.com/low-code-for-advanced-data-science

**Follow us on
social media:**



Thank you!

education@knime.com



Attachment: How to use a local update site to install extensions



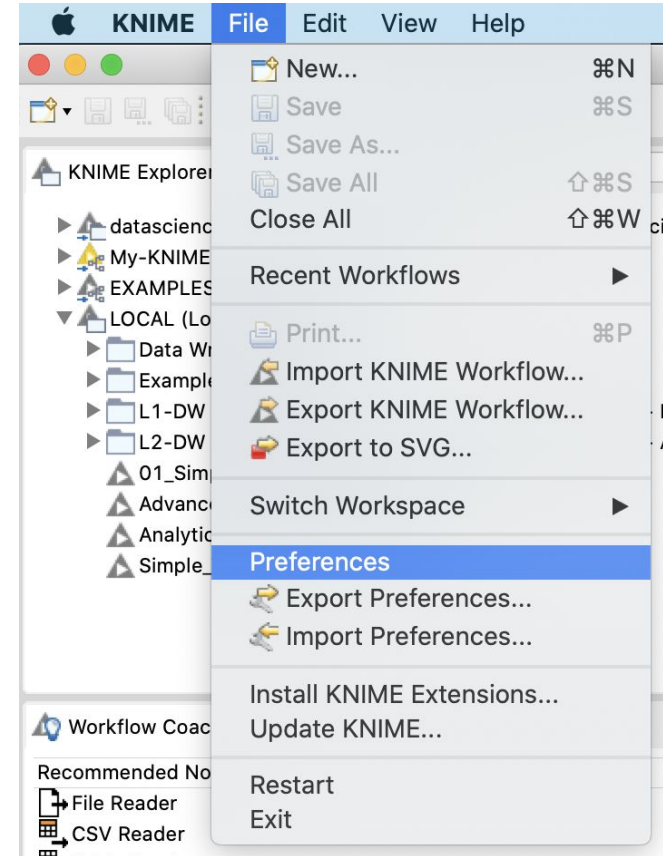
Adding a Local Update Site

- Download the update site as zip
 - [KNIME update](#) site as zip
 - [Previous versions](#) of the KNIME update site as zip
 - [Community update](#) sites as zip

Adding a Local Update Site

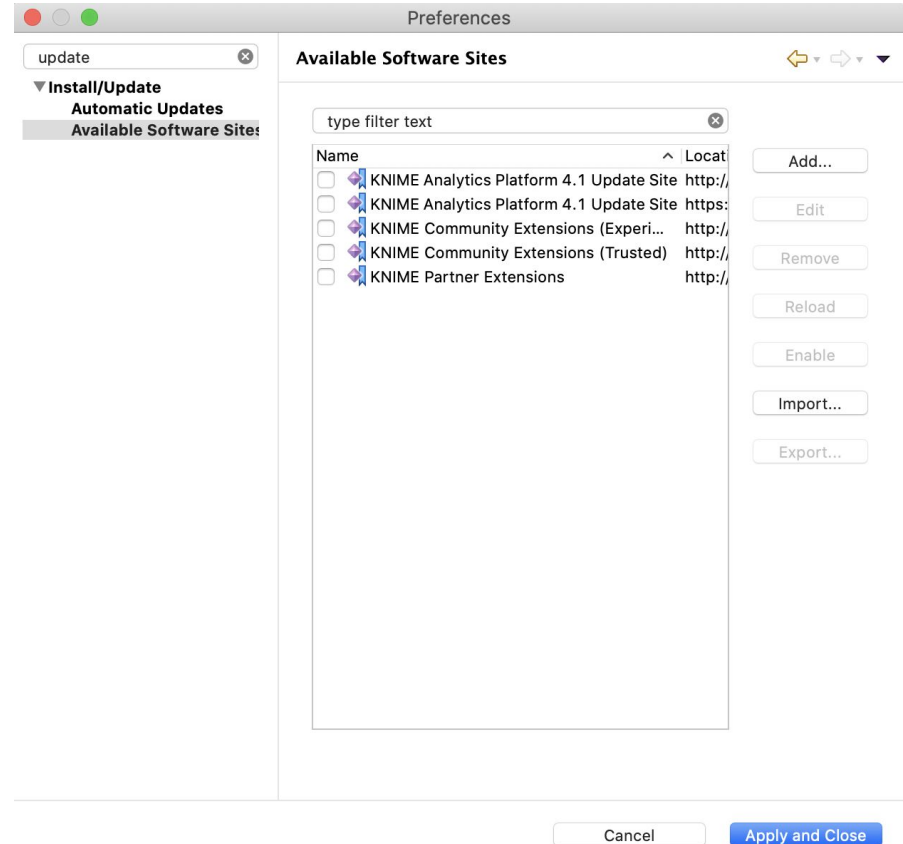
Open KNIME Analytics Platform and go to the preference page by clicking on

File -> Preferences



Adding a Local Update Site

1. Search for update (upper left search bar) and go to Available Software sites.
2. Uncheck all existing software sites.
3. Click on Add.. on the upper right.



Adding a Local Update Site

1. Define a name
2. Click on Archive and select the folder you've just downloaded
3. Click OK
4. Click Apply and Close

