

[L1-DS] KNIME Analytics Platform for Data Scientists: Basics

KNIME AG

Overview

KNIME Analytics Platform

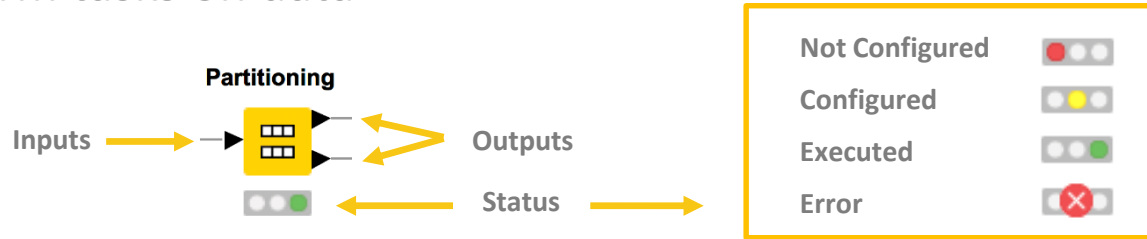


- A tool for data analysis, manipulation, visualization, and reporting
- Based on the graphical programming paradigm
- Provides a diverse array of extensions:
 - Text Mining
 - Network Mining
 - Cheminformatics
 - Many integrations, such as Java, R, Python, Weka, Keras, Plotly, H2O, etc.

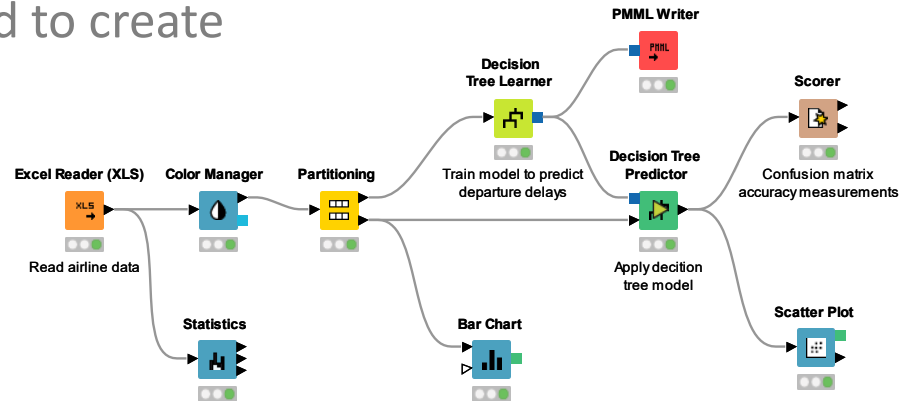


Visual KNIME Workflows

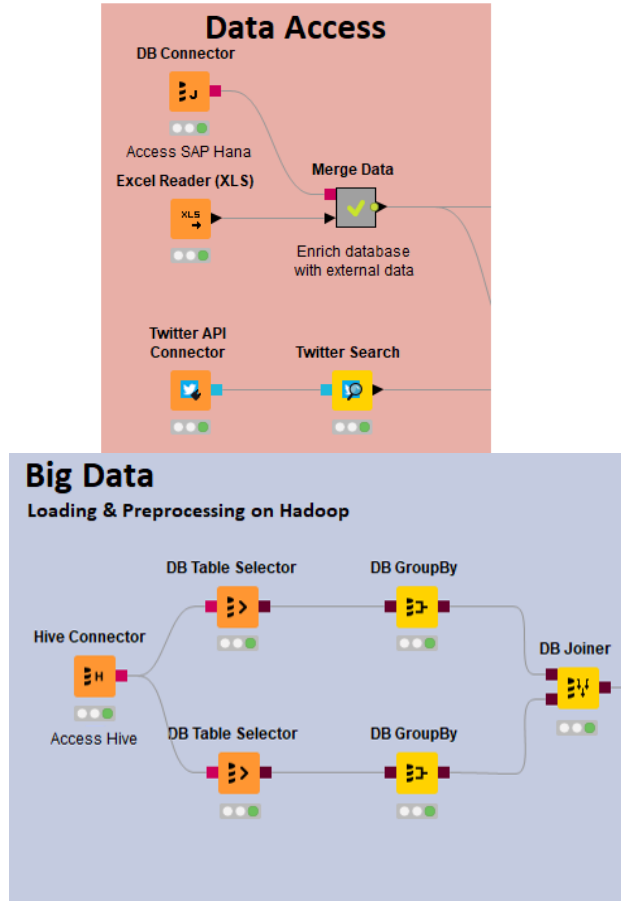
NODES perform tasks on data



Nodes are combined to create
WORKFLOWS

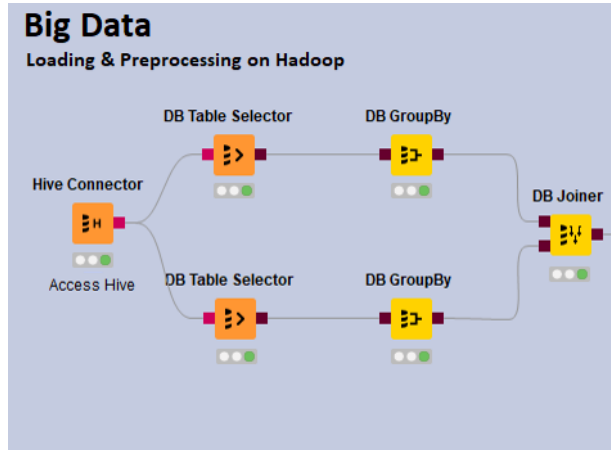


Data Access



- Databases
 - MySQL, PostgreSQL
 - any JDBC (Oracle, DB2, MS SQL Server)
- Files
 - CSV, txt
 - Excel, Word, PDF
 - SAS, SPSS
 - XML
 - PMML
 - Images, texts, networks, chem
- Web, Cloud
 - REST, Web services
 - Twitter, Google

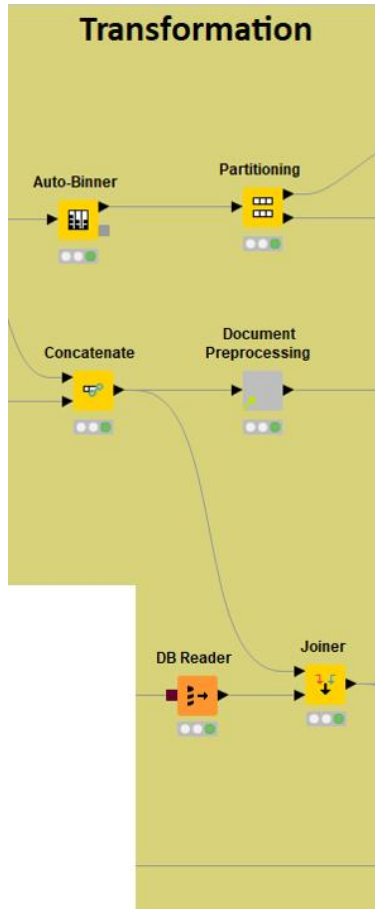
Big Data



- Spark & Databricks
- HDFS support
- Hive
- Impala
- In-database processing

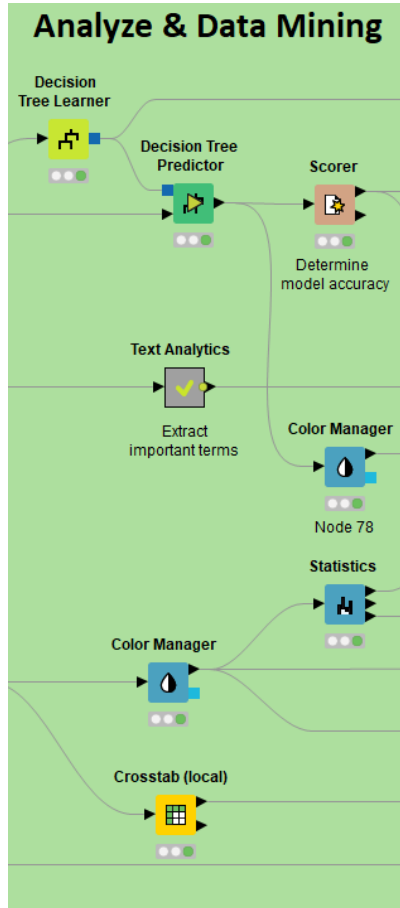


Transformation



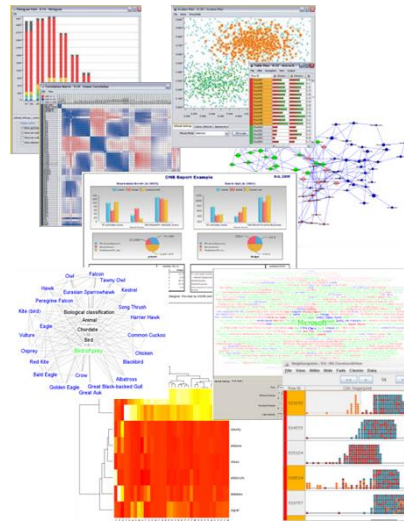
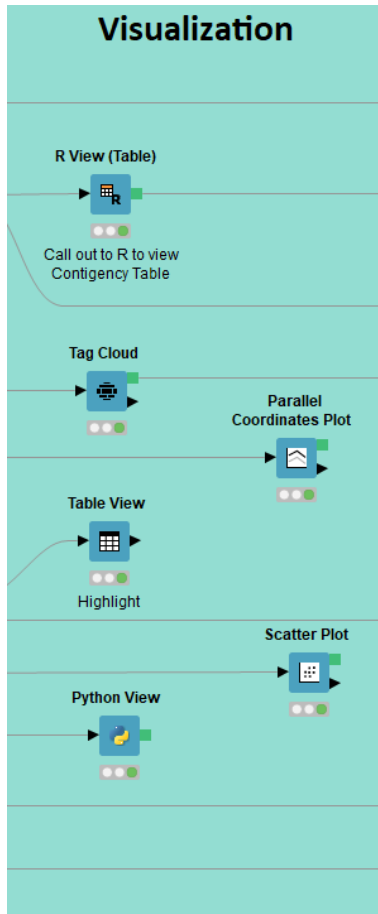
- Preprocessing
 - Row, column, matrix based
- Data blending
 - Join, concatenate, append
- Aggregation
 - Grouping, pivoting, binning
- Feature Creation and Selection

Analysis & Data Mining



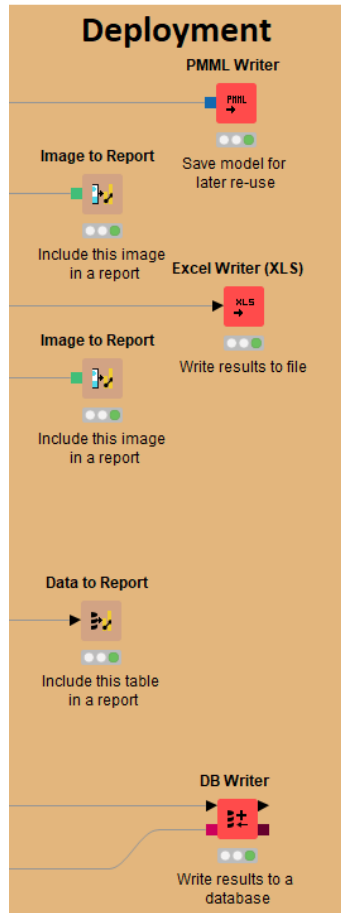
- Regression
 - Linear, logistic
- Classification
 - Decision tree, ensembles, SVM, MLP, Naïve Bayes
- Clustering
 - k-means, DBSCAN, hierarchical
- Validation
 - Cross-validation, scoring, ROC
- Deep Learning
 - Keras, DL4J
- External
 - R, Python, Weka, H2O, Keras

Visualization



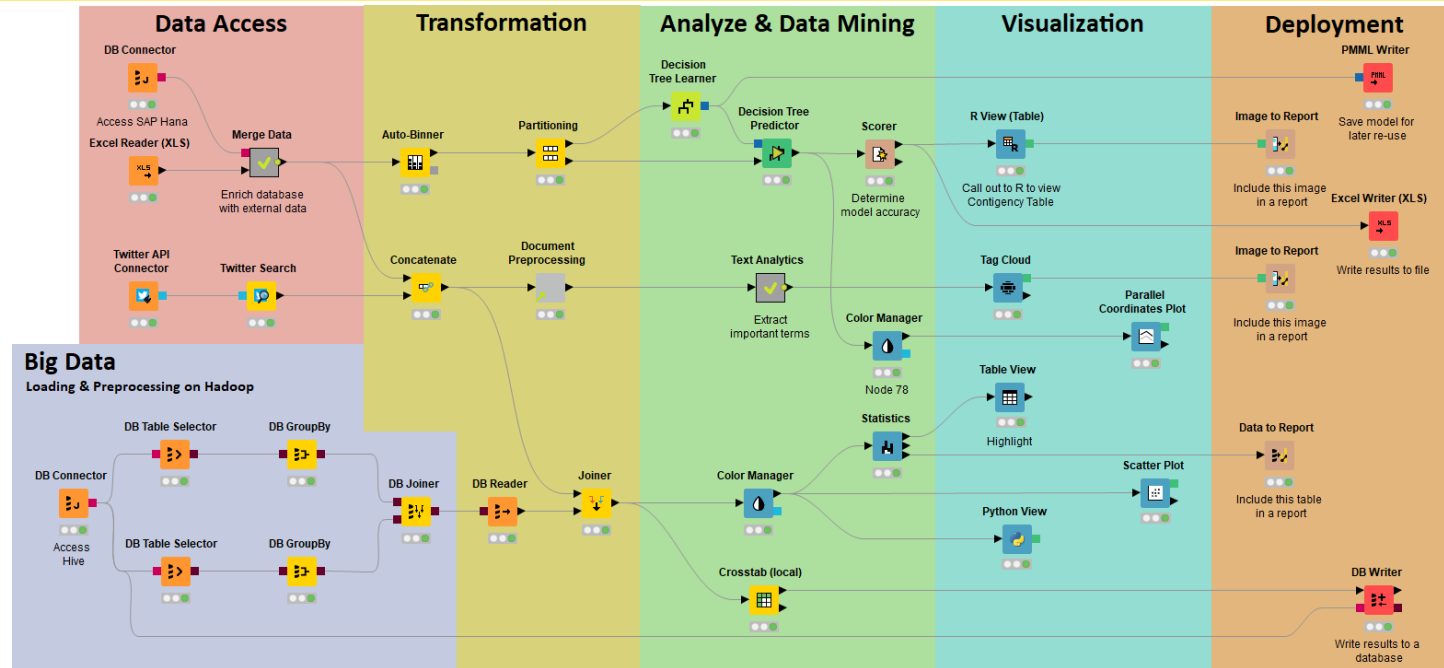
- Interactive Visualizations
- JavaScript-based nodes
 - Scatter Plot, Box Plot, Line Plot
 - Networks, ROC Curve, Decision Tree
 - Plotly Integration
 - Adding more with each release!
- Misc
 - Tag cloud, open street map, molecules
- Script-based visualizations
 - R, Python

Deployment



- Database
- Files
 - Excel, CSV, txt
 - XML
 - PMML
 - to: local, KNIME Server, SSH-, FTP-Server
- BIRT Reporting

Over 2000 Native and Embedded Nodes Included:



Data Access

MySQL, Oracle, ...
SAS, SPSS, ...
Excel, Flat, ...
Hive, Impala, ...
XML, JSON, PMML
Text, Doc, Image, ...
Web Crawlers
Industry Specific
Community / 3rd

Transformation

Row
Column
Matrix
Text, Image
Time Series
Java
Python
Community / 3rd

Analysis & Mining

Statistics
Data Mining
Machine Learning
Web Analytics
Text Mining
Network Analysis
Social Media Analysis
R, Weka, Python
Community / 3rd

Visualization

R
JFreeChart
JavaScript
Plotly
Community / 3rd

Deployment

via BIRT
PMML
XML, JSON
Databases
Excel, Flat, etc.
Text, Doc, Image
Industry Specific
Community / 3rd

Overview

- Installing KNIME Analytics Platform
- The KNIME Workspace
- The KNIME File Extensions
- The KNIME Workbench
 - Workflow editor
 - Explorer
 - Node Repository
 - Description
- Installing new features

Install KNIME Analytics Platform

- Select the KNIME version for your computer:
 - Mac
 - Windows – 32 or 64 bit
 - Linux
- Download archive and extract the file, or download installer package and run it

Windows		
KNIME Analytics Platform for Windows (installer)	32 Bit (393.38 MB)	
<i>The installer adds an icon to the desktop and suggests suitable memory settings</i>	64 Bit (396.38 MB)	
KNIME Analytics Platform for Windows (self-extracting archive)	32 Bit (396.87 MB)	
<i>The self-extracting archive only creates a folder holding the KNIME installation</i>	64 Bit (400.72 MB)	
KNIME Analytics Platform for Windows (zip archive)	32 Bit (466.11 MB)	
	64 Bit (470.07 MB)	

Linux		
KNIME Analytics Platform for Linux	64 Bit (417.21 MB)	

Mac		
KNIME Analytics Platform for Mac OSX (10.11 and above)	64 Bit (388.44 MB)	

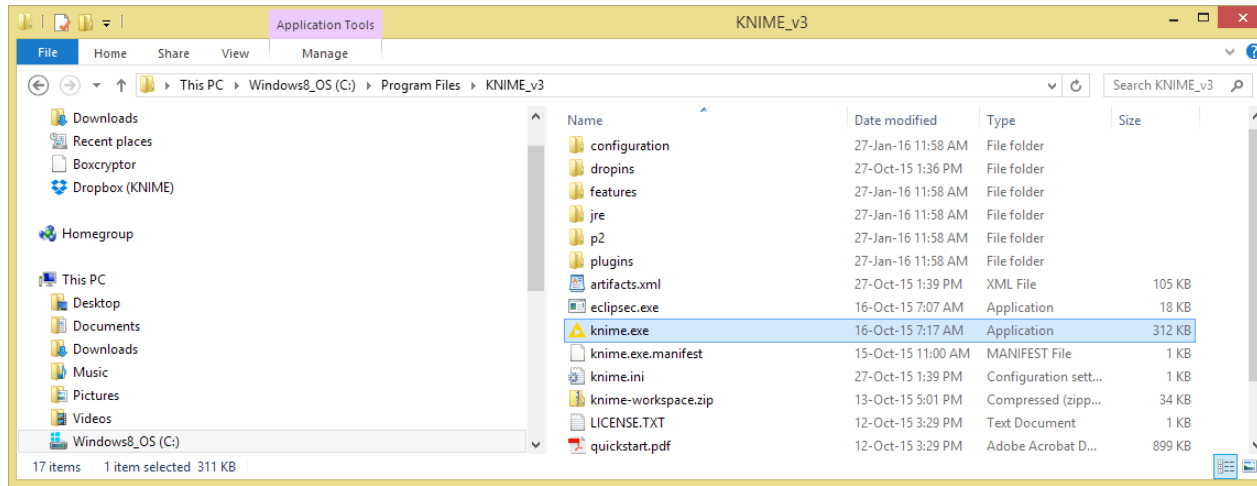
<https://www.knime.com/download-installer/2/64bit>

Start KNIME Analytics Platform

- Use the shortcut created by the installer

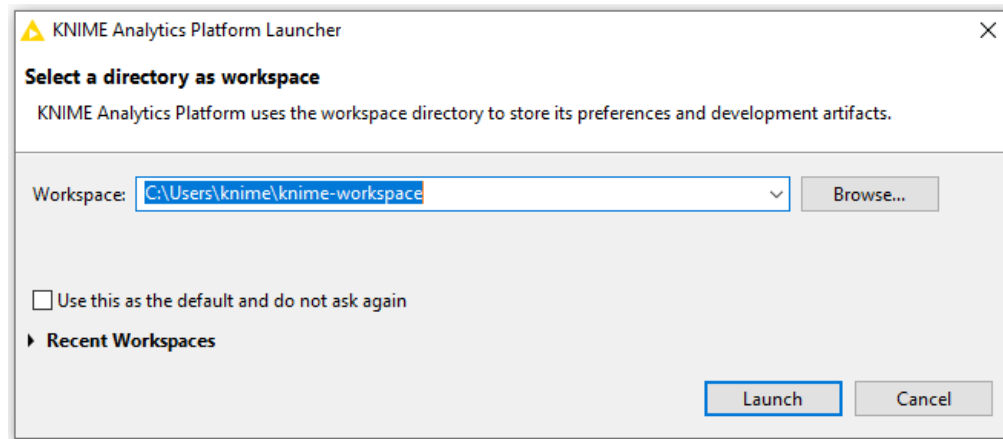


- Or go to the installation directory and launch KNIME via the knime.exe



The KNIME Workspace

- The workspace is the **folder/directory** in which workflows (and potentially data files) are stored for the current KNIME session.
- Workspaces are portable (just like KNIME)



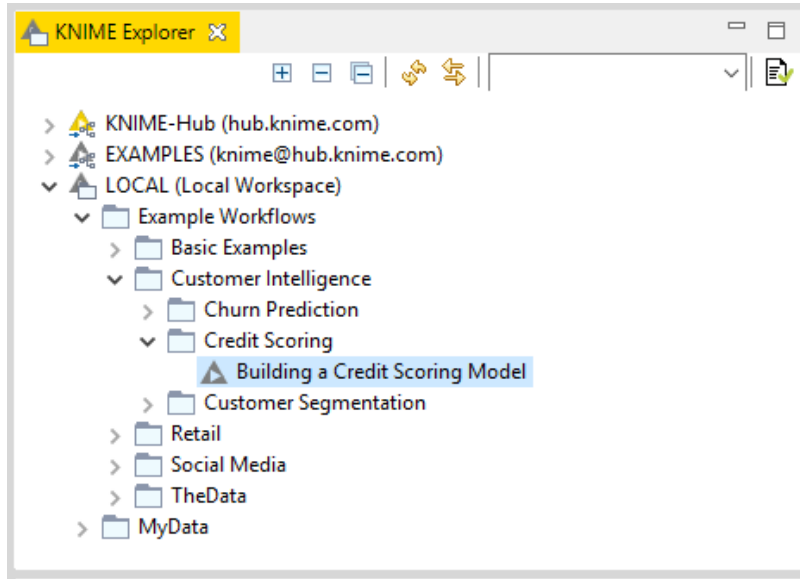
The KNIME Workbench



The screenshot displays the KNIME Analytics Platform Workbench interface. The top menu bar includes File, Edit, View, Node, and Help. The main workspace is titled "My First Workflow" and contains a workflow diagram with the following nodes: File Reader (labeled "Read sales data"), Column Filter (labeled "Include country, date and amount"), Row Filter (labeled "Exclude rows with country 'unknown'"), Stacked Area Chart (labeled "Sales over time"), and Pie/Donut Chart (labeled "Sales per country").

Yellow boxes and labels highlight the following components:

- KNIME Explorer**: Located in the top-left pane, showing the project hierarchy with folders like "My-KNIME-Hub", "EXAMPLES", "LOCAL (Local Workspace)", "Example Workflows", "My_First_Project", and "data".
- Workflow Coach**: Located in the middle-left pane, showing a list of recommended nodes including Joiner, Column Filter, Row Filter, Partitioning, GroupBy, Missing Value, and Statistics.
- Node Repository**: Located in the bottom-left pane, showing a list of node categories including IO, Manipulation, Views, Analytics, DB, and Other Data Types.
- Workflow Editor**: The central workspace where the workflow is built.
- Description**: A panel on the right side showing the description of the selected "File Reader" node.
- KNIME Hub Search**: A search bar on the right side for finding workflows and nodes.
- Outline**: A panel at the bottom showing a hierarchical view of the workflow nodes.
- Console**: A panel at the bottom showing the KNIME Console output, including the welcome message and copyright information.

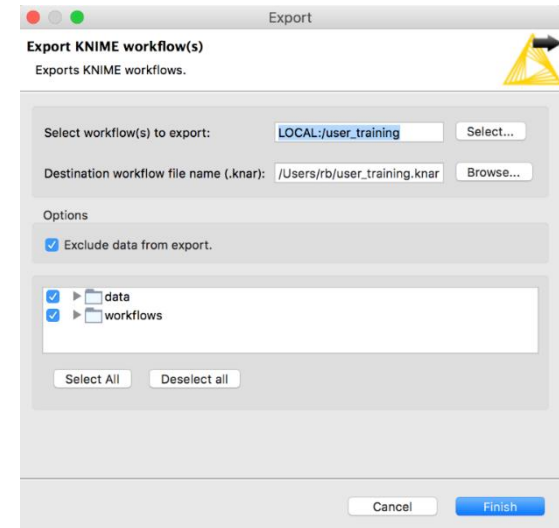
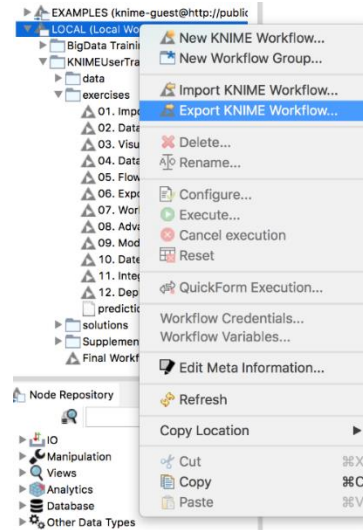
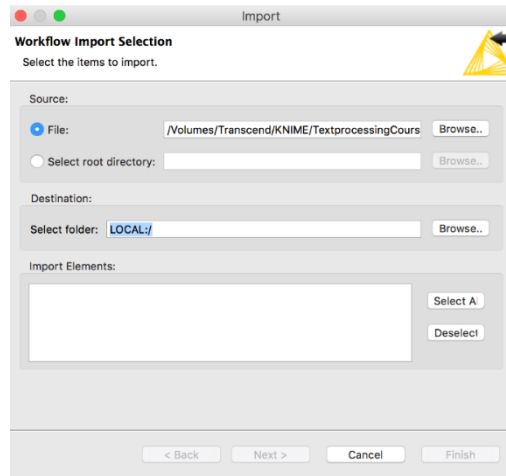
KNIME Explorer



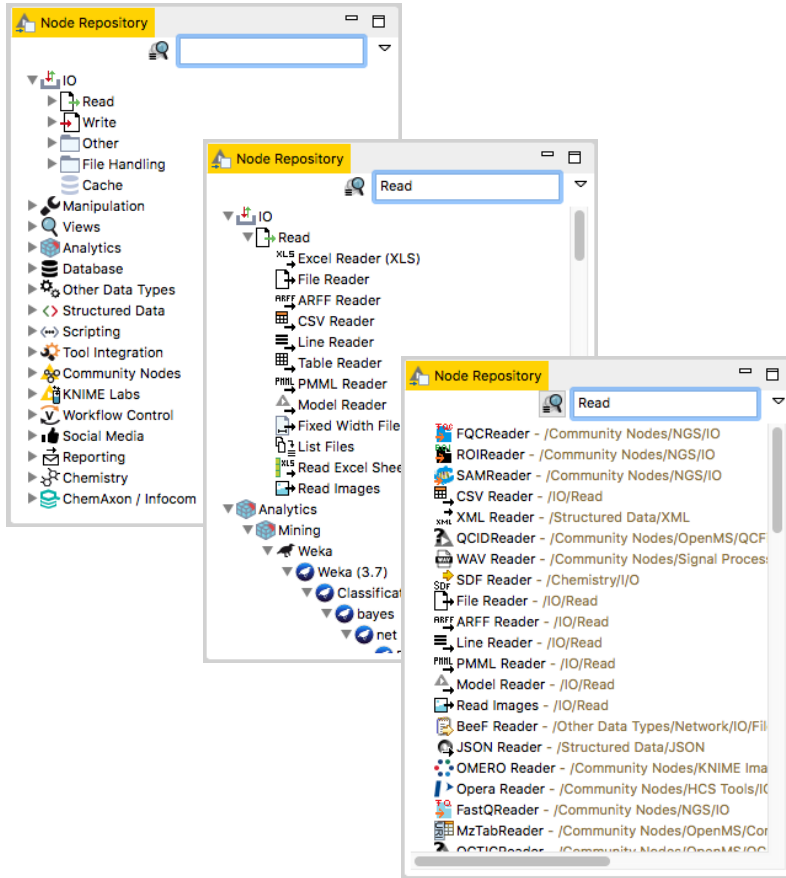
- In LOCAL you can access your own workflow projects.
- The Explorer toolbar on the top has a search box and buttons to
 -  select the workflow displayed in the active editor
 -  refresh the view
- The KNIME Explorer can contain 4 types of content:
 - Workflows
 - Workflow groups
 - Data files
 - Shared Components



Creating New Workflows, Importing and Exporting

- Right-click in KNIME Explorer to create new workflow or workflow group or to import workflow
- Right-click on workflow or workflow group to export

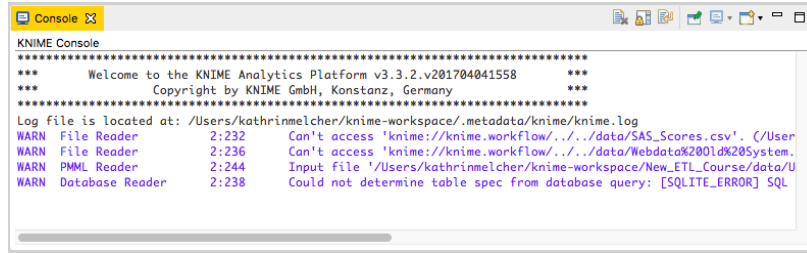


Node Repository



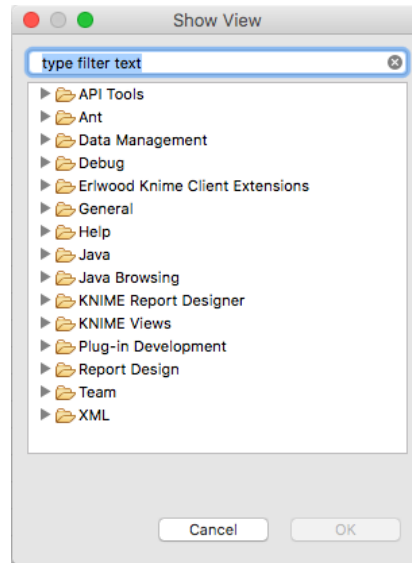
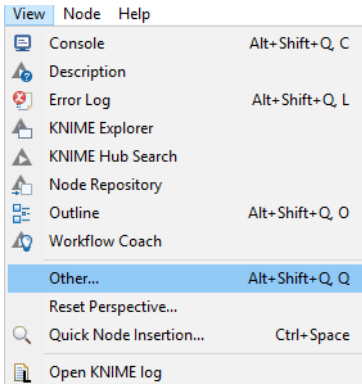
- The Node Repository lists all KNIME nodes
- The search box has 2 modes
 -  Standard Search – exact match of node name
 -  Fuzzy Search – finds the most similar node name
- Nodes can be added by drag and drop from the Node Repository to the Workflow Editor.

Console and Other Views

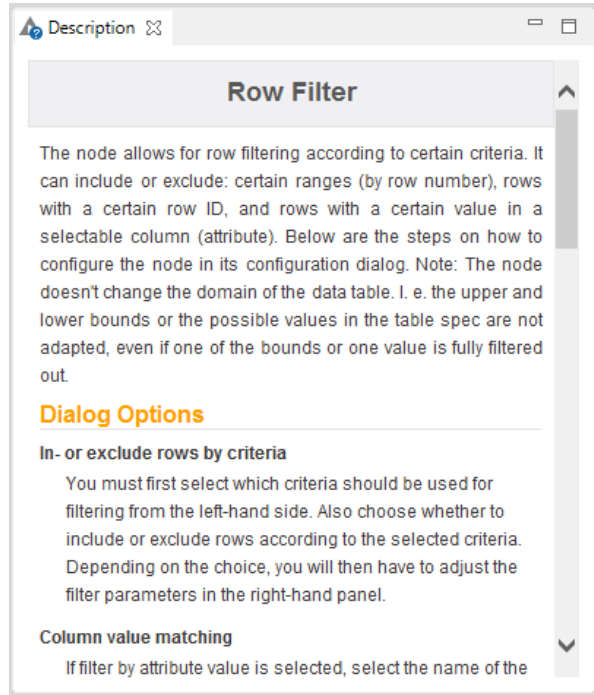


- Console view prints out error and warning messages about what is going on under the hood

- Click on View and select Other... to add different views
 - Node Monitor, Licenses, etc.

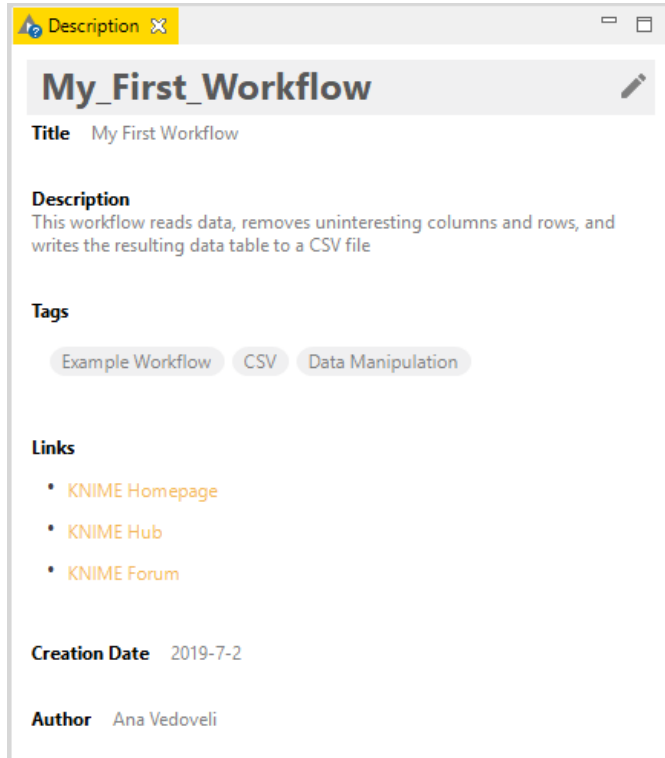


Description



- The Description window gives information about:
 - Node Functionality
 - Input & Output
 - Node Settings
 - Ports
 - References to literature

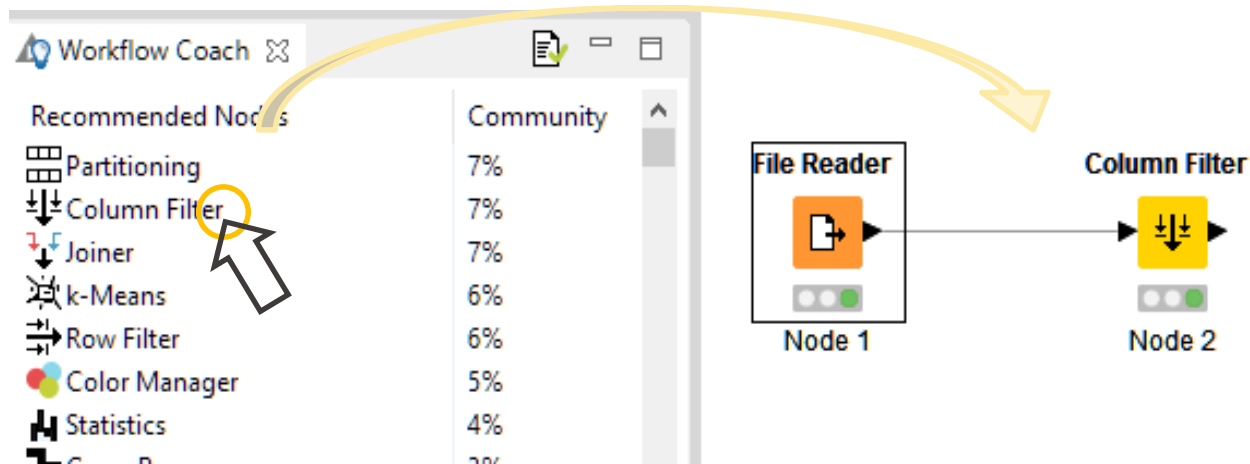
Workflow Description



- When selecting the workflow, the Description window gives information about the workflow's:
 - Title
 - Description
 - Associated Tags and Links
 - Creation Date
 - Author

Workflow Coach






- Node recommendation engine
 - Gives hints about which node use next in the workflow
 - Based on KNIME communities' usage statistics
 - Based on own KNIME workflows



Tool Bar



The buttons in the toolbar can be used for the active workflow. The most important buttons:

-  Execute selected and executable nodes (F7)
-  Execute all executable nodes
-  Execute selected nodes and open first view
-  Cancel all selected, running nodes (F9)
-  Cancel all running nodes

KNIME File Extensions

- Dedicated file extensions for Workflows and Workflow groups associated with KNIME Analytics Platform

- ***.knwf** for KNIME Workflow Files



- ***.knar** for KNIME Archive Files



More on Nodes...

A node can have 3 states:

File Reader



Not Configured:

The node is waiting for configuration or incoming data.

File Reader



Configured:

The node has been configured correctly, and can be executed.

File Reader

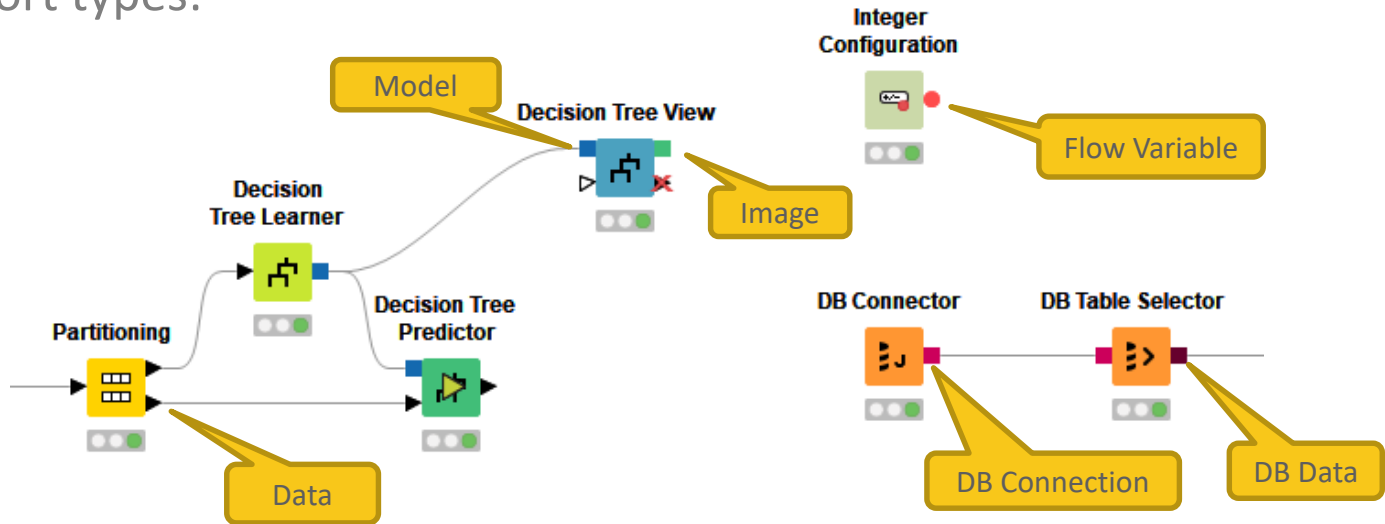


Executed:

The node has been successfully executed. Results may be viewed and used in downstream nodes.

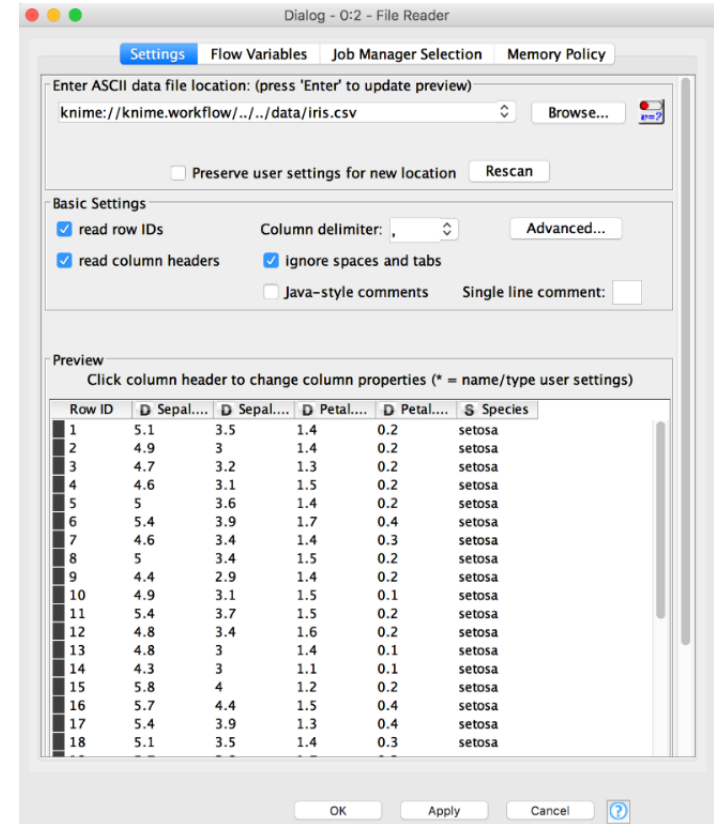
Inserting and Connecting Nodes

- Insert nodes into workspace by dragging them from Node Repository or by double-clicking in Node Repository
- Connect nodes by left-clicking output port of Node A and dragging the cursor to (matching) input port of Node B
- Common port types:



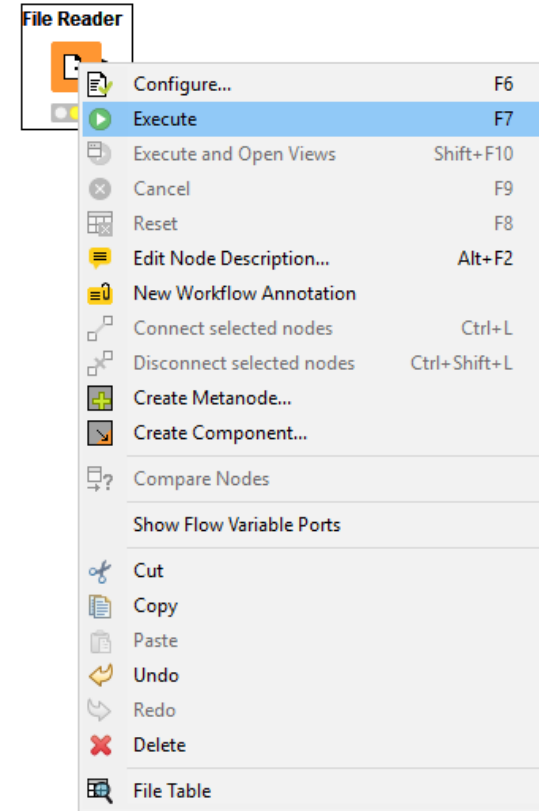
Node Configuration

- Most nodes require configuration
- To access a node configuration window:
 - Double-click the node
 - Right-click -> Configure



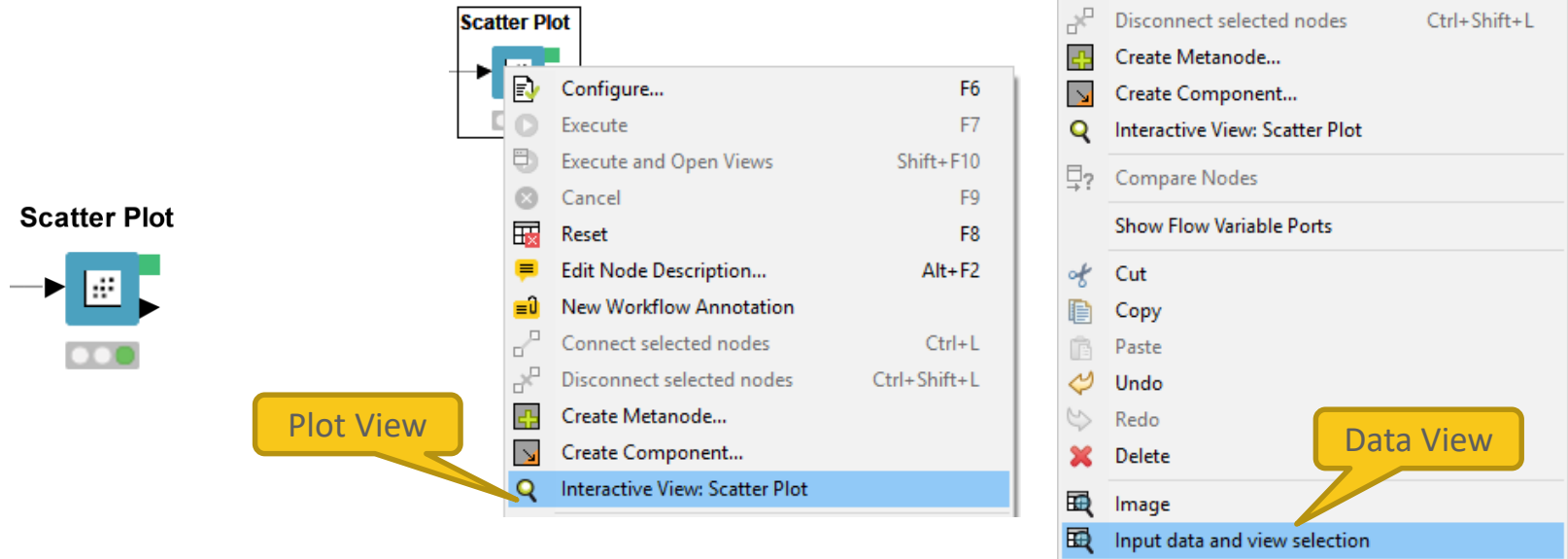
Node Execution

- Right-click node
- Select Execute in the context menu
- If execution is successful, status shows green light
- If execution encounters errors, status shows red light

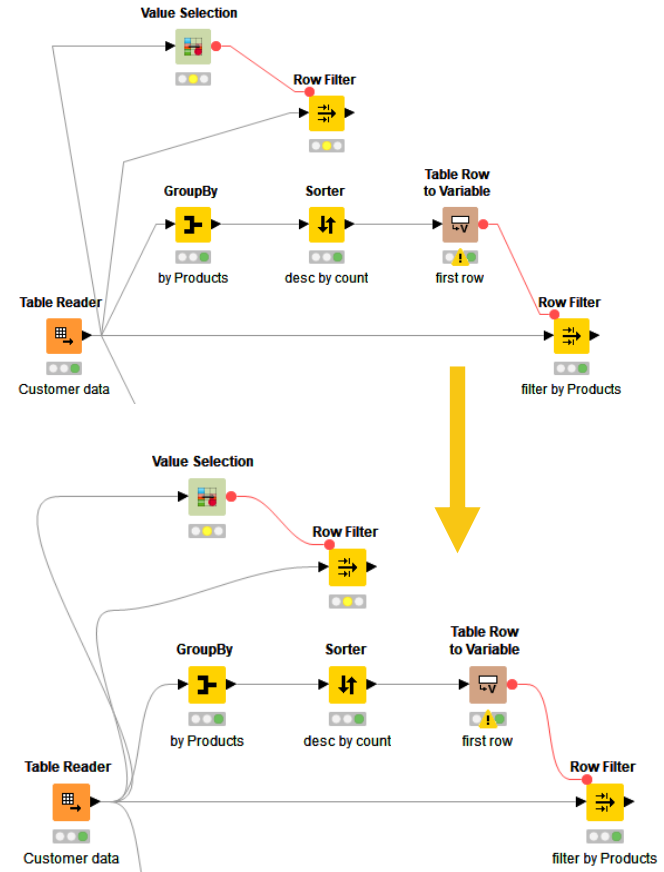
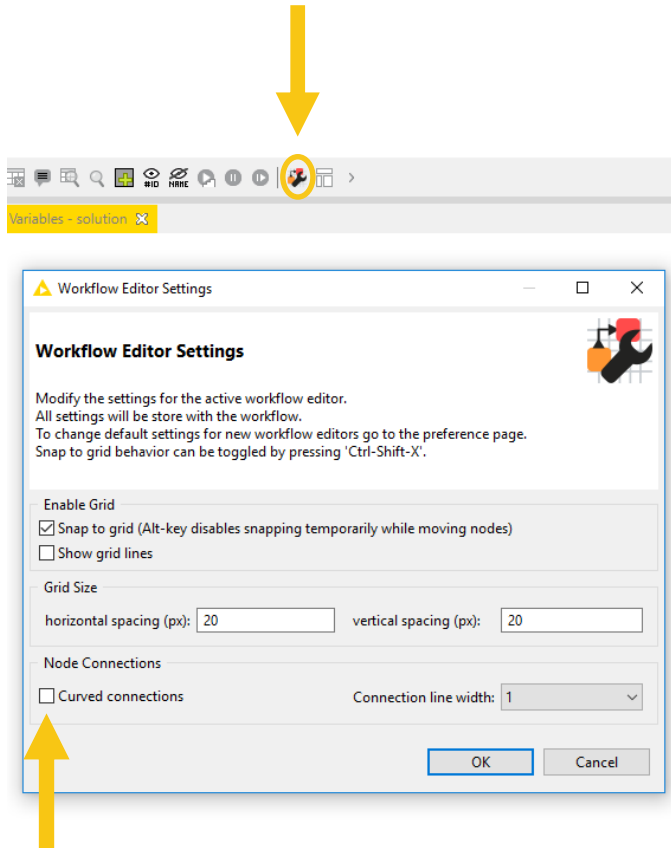


Node Views

- Right-click node
- Select Views in context menu
- Select output port to inspect execution results



Curved Connections!



Getting Started: KNIME Example Server

- Connect via KNIME Explorer to a public repository with large selection of example workflows for many, many applications
- Workflows also available on KNIME Hub

The screenshot displays the KNIME Analytics Platform interface. On the left, the **KNIME Explorer** sidebar shows a tree view of example workflows under 'My-KNIME-Hub (hub.knime.com)' and 'LOCAL (Local Workspace)'. The central workspace shows a workflow titled 'Simple Preprocessing Example' with a yellow box highlighting its purpose: 'Shows the use of different filter nodes.' The workflow consists of the following nodes: **File Reader** (input), **Row Filter** (keep rows where 'native-country' is missing), **Column Filter** (selects columns for further processing), **Reference Column Filter** (used by the Column Filter), **Numeric Binner** (replace numeric column age with an attribute column), **Nominal Value Row Filter** (filter by this new attribute), **Reference Row Filter** (used by the Row Filter), **Row Filter** (filter example with a regular expression), and **Concatenate** (combines the filtered data). On the right, a separate **KNIME Explorer** window shows a list of example workflows, including 'My-KNIME-Hub (hub.knime.com)', 'EXAMPLES (knime@hub.knime.com)', and 'LOCAL (Local Workspace)'. A tooltip indicates 'Double-click to see the examples'.

Sharing Workflows

How to use the KNIME Hub



KNIME Hub



Welcome to the KNIME Hub

The place to find and collaborate on KNIME workflows and nodes. Here you can find solutions for your data science questions.

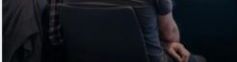
Search workflows, nodes and more...

How to Getting Started

From downloading through to building your first workflow



Forum Get help from our community and help others



Blog Intelligently Automating Machine Learning



Search workflows, nodes and more...

More KNIME

Training a Churn Predictor



This workflow is an example of how to build a basic PAMM model for a churn prediction using a Decision Tree algorithm.

External Resources

Churn Prediction



Search workflows, nodes and more...

More KNIME

Scatter Plot



A scatter plot using a JavaScript based charting library. The view can be accessed either via the "Interactive view" action on the executed node or in KNIME Server web portal page.



The configuration of the node lets you choose the size of a sample to display and to enable certain controls, which are then available in the view. This includes the ability to choose different columns for x and y or the possibility to set a title. Enabling or disabling these controls via the configuration dialog might not seem useful at first glance but has benefits when used in a web portal/external execution where the end user has no access to the workflow itself.

Since missing values as well as NaN (not a number) or infinite values cannot be displayed in the view, they will be omitted with a corresponding warning message.

Additionally a static SVG image can be rendered, which is then made available at the first output port.

Note, this node is currently under development. Future versions of the node might have more or changed functionality.

Ports Options Views

Input Ports
Type Data
Data table with data to display

Output Ports
Type Image
SVG image rendered by the JavaScript implementation of the scatter plot.
Type Data
Data table containing the input data appended with a column, that represents the selection made in the scatter plot view.

Extension

This node is part of the extension
KNIME JavaScript Views V1.0.0

Short Link

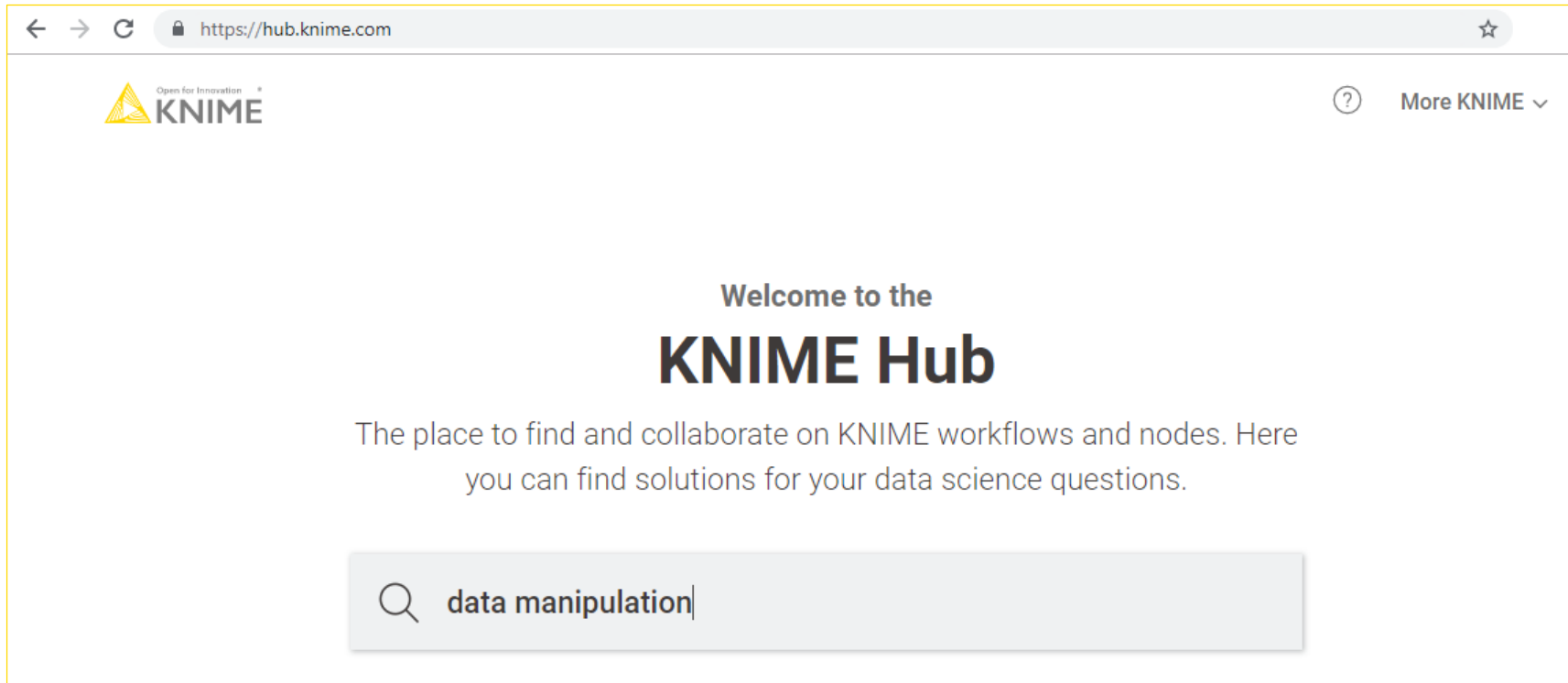
https://hub.knime.com/extension/26



Drag node into KNIME Analysis Platform

A place to share knowledge about
Workflows and Nodes
<https://hub.knime.com>

The KNIME Hub



Searching Nodes and Workflows

The screenshot shows the KNIME search interface with the search term "data manipulation". The results page displays 3140 results. The search bar at the top includes the KNIME logo, the search term, a close button, and a "More KNIME" link. Below the search bar, there are tabs for "All", "Nodes", and "Workflows". The "Nodes" tab is selected, showing a list of search results. The first result is "String Manipulation", which is a node. The second result, "String Manipulation, Math Formula and Rule Engine Example", is a workflow and is highlighted with a dashed yellow border. The third result is "String Manipulation (Variable)", which is also a node. Each result card includes a small icon, the title, a description, a list of tags, a breadcrumb trail, and a thumbnail image of the node or workflow.

3140 results


All Nodes Workflows

String Manipulation
streamable
Manipulates strings like search and replace, capitalize or remove leading and trailing white spaces. Examples: To remove leading and trailing blanks from a column with name c0 you would use the expres...
Manipulation > Column > Convert & Replace
Manipulator

String Manipulation, Math Formula and Rule Engine Example
ETL data manipulation string manipulation strings numbers.math math formula data transformation data wrangling rules rule engine
This workflow shows three different data manipulation operations, namely: - creating three categories of people based on their weekly work hours with the Rule Engine node - rounding up people's age to...
Users > knime > Examples > 02_ETL_Data_Manipulation > 04_Transformation


String Manipulation (Variable)
Manipulates or defines values of variables like search and replace, capitalize or remove leading and trailing white spaces. Examples: To remove leading and trailing blanks from a variable with name c0...
Workflow Control > Variables
Manipulator

Opening a Workflow from the Hub

 Open for Innovation

Search workflows, nodes and more...

? More KNIME ▾

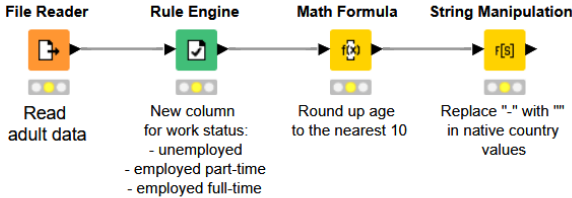


String Manipulation, Math Formula and Rule Engine Example

String Manipulation, Math Formula and Rule Engine Example

This workflow shows three different data manipulation operations, namely:


- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with "" characters in the native country column



```
graph LR; A[File Reader] --> B[Rule Engine]; B --> C[Math Formula]; C --> D[String Manipulation]
```


This workflow shows three different data manipulation operations, namely:

- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with "" characters in the native country column



Maarit
KNIME Team Member

Basic
Silver Anniversary
First Share 5 more

hosted by
 KNIME

Open workflow
or [download workflow](#)

By downloading the workflow, you agree to our [terms and conditions](#).

CC-BY-4.0

Short Link
<https://knl.me/w/pyg3VYLc9BL4sJPd>

Open Workflow in KNIME Analytics Platform

The screenshot displays the KNIME Analytics Platform interface. The top menu bar includes File, Edit, View, Node, and Help. The left sidebar contains the KNIME Explorer and Node Repository. The main workspace shows a workflow titled "3: 02_StringManipulation_MathFormula_RuleEngine".

KNIME Explorer:

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)

Node Repository:

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- KNIME Labs
- Workflow Control
- Workflow Abstraction
- Social Media
- Reporting
- Chemistry

Workflow Details:

This is a temporary copy of "knime://EXAMPLES/Users/knime/Examples/02_ETL_Data_Manipulation/04_Transforma..."

String Manipulation, Math Formula and Rule Engine Example

This workflow shows three different data manipulation operations, namely:

- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with " " characters in the native country column

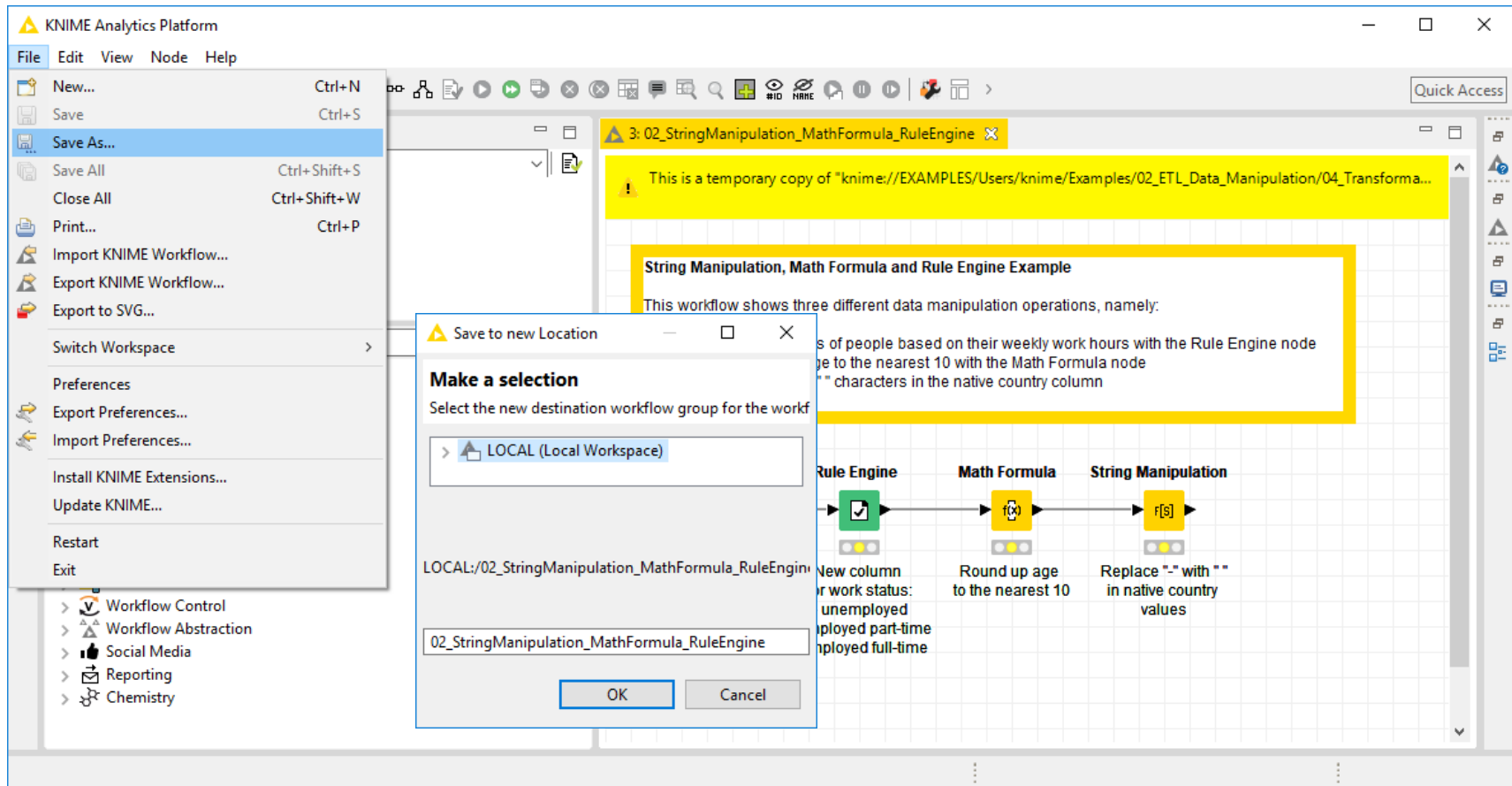
Workflow Diagram:

```
graph LR; A[File Reader] --> B[Rule Engine]; B --> C[Math Formula]; C --> D[String Manipulation]
```

Node Descriptions:

- File Reader:** Read adult data
- Rule Engine:** New column for work status:
 - unemployed
 - employed part-time
 - employed full-time
- Math Formula:** Round up age to the nearest 10
- String Manipulation:** Replace "-" with " " in native country values


Saving the Workflow



Edit the Workflow

Row Filter

The node allows for row filtering a criteria. It can include or exclude: row number), rows with a certain a certain value in a selectable col are the steps on how to configure configuration dialog. Note: the domain of the data table. I. e. the upper and lower possible values in the table spec are not adapted, even or one value is fully filtered out.



Manipulator

KNIME Analytics Platform

File Edit View Node Help

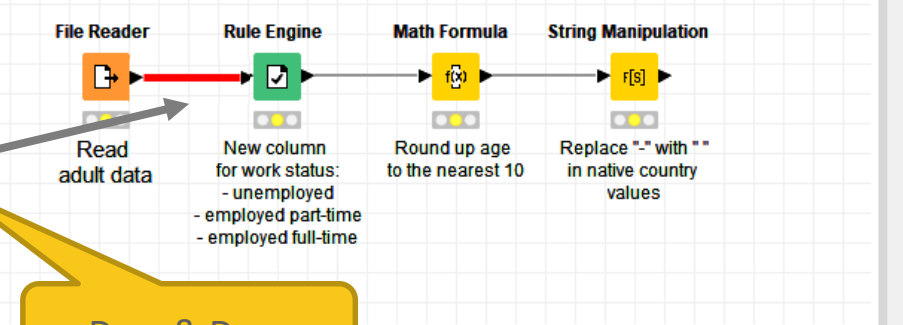
100%

*0: 02_StringManipulation_MathFormula_RuleEngine

String Manipulation, Math Formula and Rule Engine Example

This workflow shows three different data manipulation operations, namely:

- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with "" characters in the native country column



```
graph LR; A[File Reader] --> B[Rule Engine]; B --> C[Math Formula]; C --> D[String Manipulation]
```

File Reader
Read adult data

Rule Engine
New column for work status:
- unemployed
- employed part-time
- employed full-time

Math Formula
Round up age to the nearest 10

String Manipulation
Replace "-" with "" in native country values

Drag & Drop

Sharing the Workflow

The screenshot displays the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' pane shows a tree view with 'My-KNIME-Hub (hub.knime.com)' selected. A yellow callout bubble points to the 'Save' icon in the top toolbar, labeled '1. Save your Edits'. Another yellow callout bubble points to the 'My-KNIME-Hub' entry, labeled '2. Connect to KNIME Hub'. Below the Explorer, the 'Node Repository' pane lists various node categories like IO, Manipulation, Views, etc. The main workspace shows a workflow titled '*0: 02_StringManipulation_MathFormula_RuleEngine'. A yellow box highlights a text area describing the workflow: 'This workflow shows three different data manipulation operations, namely: - creating three categories of people based on their weekly work hours with the Rule Engine node - rounding up people's age to the nearest 10 with the Math Formula node - replacing hyphens with "" characters in the native country column'. Below this text, a diagram illustrates the workflow steps: File Reader (Read adult data) -> Row Filter (filter data) -> Rule Engine (New column for work status: - unemployed, - employed part-time, - employed full-time) -> Math Formula (Round up age to the nearest 10) -> String Manipulation (Replace "-" with "" in native country values).

1. Save your Edits

2. Connect to KNIME Hub

Rule Engine Example

This workflow shows three different data manipulation operations, namely:

- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with "" characters in the native country column

Workflow Diagram:

```
graph LR; A[File Reader] --> B[Row Filter]; B --> C[Rule Engine]; C --> D[Math Formula]; D --> E[String Manipulation]
```

File Reader: Read adult data

Row Filter: filter data

Rule Engine: New column for work status:
- unemployed
- employed part-time
- employed full-time

Math Formula: Round up age to the nearest 10

String Manipulation: Replace "-" with "" in native country values

Log in the Hub

The image displays the KNIME Analytics Platform interface on the left and a web browser window on the right, illustrating the login process.

KNIME Analytics Platform:

- The **KNIME Explorer** pane shows a tree structure with **My-KNIME-Hub (hub.knime.com)** selected. A tooltip indicates: "Double click to connect to KNIME Hub".
- The **Node Repository** pane lists various nodes and categories, including **IO**, **Machine Learning**, **Visualization**, **Analytics**, **Database**, **Other Data Types**, **Structured Data**, **Scripting**, **Tools & Services**, **KNIME Labs**, **Workflow Control**, **Workflow Abstraction**, **Social Media**, **Reporting**, and **Chemistry**.
- A dialog box titled "Login to 'My-KNIME-Hub': A new tab has been opened in your browser." is displayed, with **cancel** and **retry** buttons.

Web Browser Window:

- The browser shows the **Log in | KNIME** page at the URL <https://www.knime.com/user/login?destination=oauth2/au...>.
- The page features the **KNIME** logo and the tagline "Open for Innovation".
- The **Log in** section includes input fields for email (containing "paolotamag") and password (masked with dots).
- A yellow speech bubble points to the email field with the text: **KNIME Forum Account Credentials**.
- Below the password field is a link for [Forgotten password?](#).
- A yellow **Log in** button is present.
- At the bottom, the **Create Account** link is circled in yellow.

Publish your Workflow

The screenshot displays the KNIME interface with three main components: a metadata editor, a file explorer, and a workflow canvas.

1. Edit Metadata

The **02_StringManipulation_MathFormula_RuleEngine** workflow is selected in the **KNIME Explorer**. The **Description** tab is active, showing the workflow's title and description. The **Tags** section lists **ETL**, **data manipulation**, **numbers.math**, **math form**, **data wrangling**, and **rules**. The **Links** section shows the URL [e.g. https://www.knime.com](https://www.knime.com), the title [e.g. Outlier detection](#), and the type **Website**.

2. Drag & Drop

The workflow is published to the **My-KNIME-Hub** (maatit@hub.knime.com) under the **Public** visibility. The workflow is titled **02_StringManipulation_MathFormula_RuleEngine**.

String Manipulation, Math Formula and Rule Engine Example

This workflow shows three different data manipulation operations, namely:

- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with " " characters in the native country column

The workflow diagram shows the following steps:

- File Reader**: Read adult data
- Row Filter**
- Rule Engine**: New column for work status:
 - unemployed
 - employed part-time
 - employed full-time
- Math Formula**: Round up age to the nearest 10
- String Manipulation**: Replace "-" with " " in native country values

Open your Workflow in the Hub

The screenshot shows the KNIME Explorer interface. On the left, a tree view displays the project structure under 'My-KNIME-Hub'. The workflow '02_StringManipulation_MathFormula_RuleEngine' is selected. A context menu is open over this workflow, with the 'Open' option highlighted. A sub-menu is visible, showing 'as Local Copy' and 'in KNIME Hub' options. The 'in KNIME Hub' option is selected. On the right, the workflow canvas is visible, showing a sequence of nodes: File Reader, Row Filter, Rule Engine, Math Formula, and String Manipulation. A yellow box highlights the workflow title and a description of its operations.

String Manipulation, Math Formula and Rule Engine Example

This workflow shows three different data manipulation operations, namely:

- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with " " characters in the native country column

File Reader
Read adult data

Row Filter

Rule Engine
New column for work status:
- unemployed
- employed part-time
- employed full-time

Math Formula
Round up age to the nearest 10

String Manipulation
Replace "-" with " " in native country values

Open your Workflow in the Hub

The screenshot shows a web browser window displaying the KNIME Hub interface. The address bar shows the URL: `hub.knime.com/maarit/spaces/Public/latest/02_StringManipulation_MathFormula_RuleEngine`. The breadcrumb navigation path is: **KNIME Hub** > **maarit** > **Spaces** > **Public** > **02_StringManipulation_MathFormula_RuleEngine**. The main heading is **String Manipulation, Math Formula and Rule Engine Example**. Below the heading, a description states: "This workflow shows three different data manipulation operations, namely:" followed by a list:

- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with "" characters in the native country column

 The workflow diagram consists of five nodes: **File Reader** (Read adult data), **Row Filter**, **Rule Engine** (New column for work status: - unemployed, - employed part-time, - employed full-time), **Math Formula** (Round up age to the nearest 10), and **String Manipulation** (Replace "-" with "" in native country values). On the right side, there is a user profile for **Maarit**, a **Download workflow** button, a license of **CC-BY-4.0**, and a **Short link** section containing the URL `https://knime.me/wjft1y0PpalmtC8Z45`.

Hot Keys (for Future Reference)

Task	Hot key	Description
Node Configuration	F6	opens the configuration window of the selected node
Node Execution	F7	executes selected configured nodes
	Shift + F7	executes all configured nodes
	Shift + F10	executes all configured nodes and opens all views
	F9	cancels selected running nodes
	Shift + F9	cancels all running nodes
Node Connections	Ctrl + L	connects selected nodes
	Ctrl + Shift + L	disconnects selected nodes
Move Nodes and Annotations	Ctrl + Shift + Arrow	moves the selected node in the arrow direction
	Ctrl + Shift + PgUp/PgDown	moves the selected annotation in the front or in the back of all overlapping annotations
Workflow Operations	F8	resets selected nodes
	Ctrl + S	saves the workflow
	Ctrl + Shift + S	saves all open workflows
	Ctrl + Shift + W	closes all open workflows
Metanode	Shift + F12	opens metanode wizard

Stay connected with KNIME



Blog: knime.com/blog



Forum: forum.knime.com



KNIME Hub:
hub.knime.com



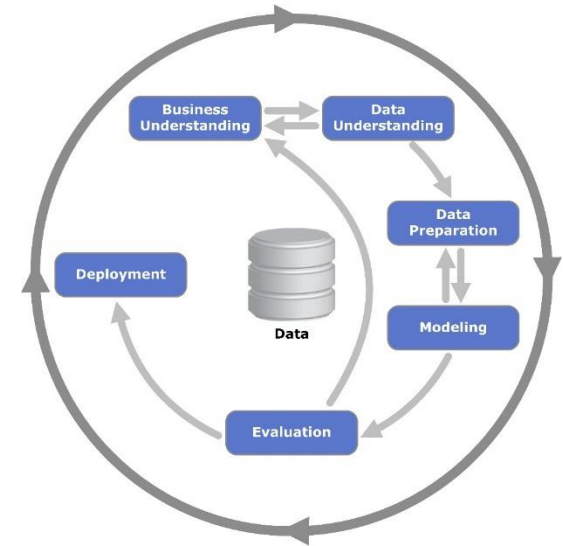
KNIME E-Learning Course:
www.knime.com/e-learning-course

Follow us on social
media:

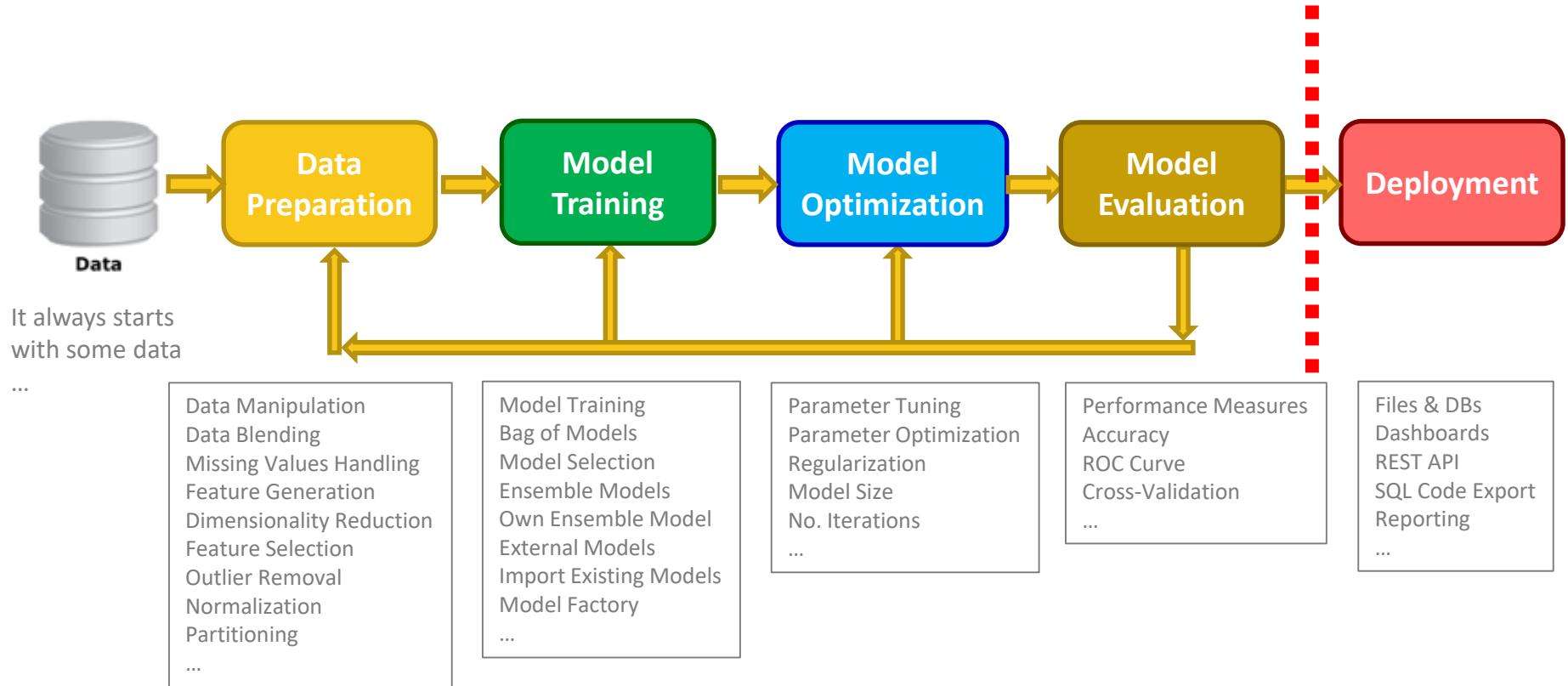


Today's Example: Churn Prediction

- Build a data science application step by step
- Each section of the course has an associated workflow with exercises
- The exercises complete the steps in the CRISP-DM cycle

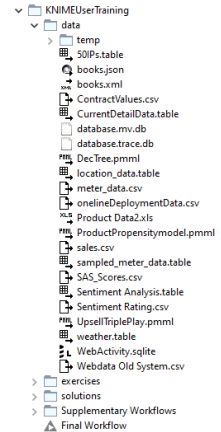


Today's Example: Churn Prediction

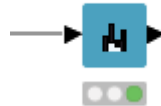


The Data

- The data files used in the exercises are available in the “data” folder: data files in different file formats, web-based data, data on a database, etc.
- For churn prediction, customer data are blended from different sources
- The Data Explorer node is helpful in inspecting data



Data Explorer



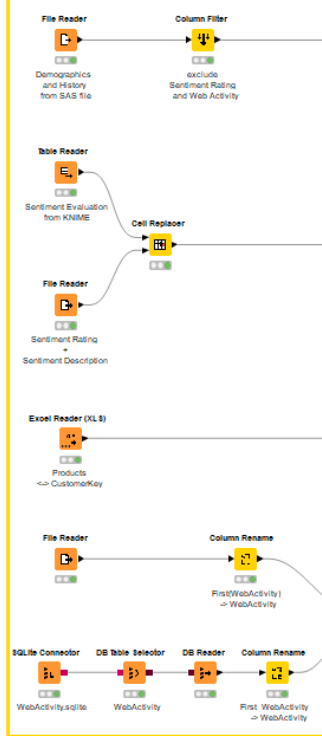
Column	Exclude Column	Minimum	Maximum	Mean	Standard Deviation
CustomerKey	<input type="checkbox"/>	11000	27336	19281.750	5319.909
WebActivity	<input type="checkbox"/>	0	5	1.000	1.524
SentimentRating	<input type="checkbox"/>	0	5	1.846	1.619
EstimatedYearlyIncome	<input type="checkbox"/>	10000	170000	57066.921	32242.624
NumberOfContracts	<input type="checkbox"/>	0	4	1.493	1.145
Age	<input type="checkbox"/>	29	100	48.288	11.382
Target	<input type="checkbox"/>	0	1	0.489	0.500

Today's Example: Churn Prediction

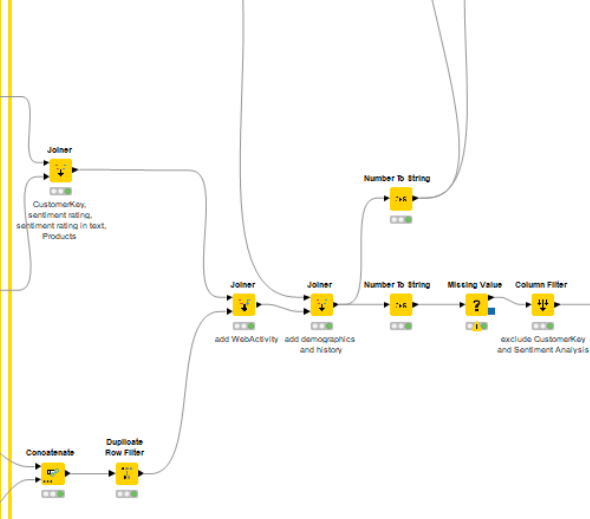
Final Workflow from the KNIME User Training

...and putting all those parts together, you get this final workflow:

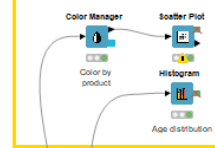
Data Reading



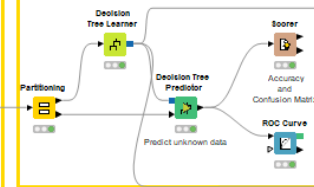
Data Manipulation and Aggregation



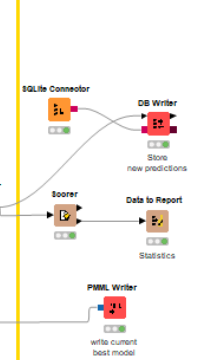
Visualization



Training Predictive Models



Data Export and Reporting



Importing Data

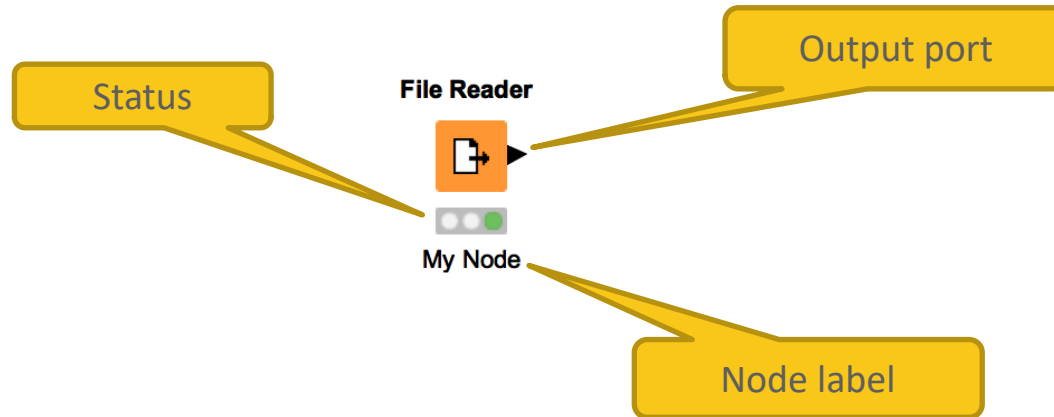
Accessing Files and Databases



Data Source Nodes

Typically characterized by:

- Orange color
- No input ports, 1-2 output ports



New Node: File Reader

Workhorse of the KNIME Source nodes

- Reads all text based files (e.g. csv, txt, etc.)
- Many advanced features allow it to read most 'weird' files
 - Short lines, inline comments, headers and special encoding

File Reader



My Node

YouTube KNIME TV Channel video:

<https://youtu.be/flaHQw-Qhlg>

File Reader Configuration

The screenshot shows the 'File Reader (My Node)' configuration dialog. It has tabs for 'Settings', 'Flow Variables', 'Job Manager Selection', and 'Memory Policy'. The 'Settings' tab is active, showing fields for 'Enter ASCII data file location' (with a text input and a 'Browse...' button), a 'Preserve user settings for new location' checkbox, and a 'Rescan' button. Below this is the 'Basic Settings' section with checkboxes for 'read row IDs', 'read column headers', 'ignore spaces and tabs', and 'Java-style comments', along with a 'Column delimiter' dropdown and an 'Advanced...' button. The 'Preview' section at the bottom shows a table of data with columns 'Row ID', 'D Sepal...', 'D Sepal...', 'D Petal...', 'D Petal...', and 'S Species'. A callout labeled 'File path' points to the file location input. A callout labeled 'Basic Settings' points to the 'read row IDs' checkbox. A callout labeled 'Advanced Settings' points to the 'Advanced...' button. A callout labeled 'Preview' points to the data table. A callout labeled 'Help Button' points to the question mark icon at the bottom right.

File path

Basic Settings

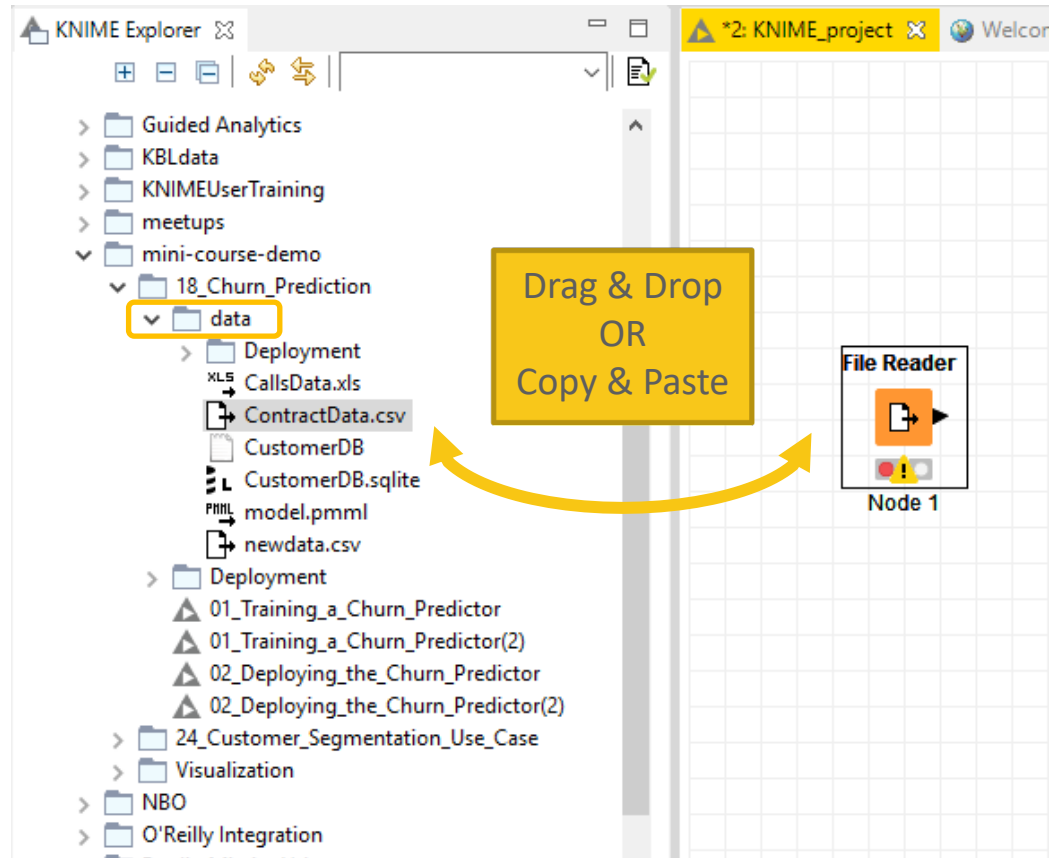
Advanced Settings

Preview

Help Button

Row ID	D Sepal...	D Sepal...	D Petal...	D Petal...	S Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3	1.4	0.1	setosa
14	4.3	3	1.1	0.1	setosa
15	5.8	4	1.2	0.2	setosa
16	5.7	4.4	1.5	0.4	setosa
17	5.4	3.9	1.3	0.4	setosa
18	5.1	3.5	1.4	0.3	setosa

Alternative Faster Way ...



Filenames and the knime:// Protocol

Absolute URL

Input location

knime://LOCAL/KNIMEUserTraining/data/Sentiment%20Analysis.table

Browse...

Mountpoint-relative URL

Input location

knime://knime.mountpoint/KNIMEUserTraining/data/Sentiment%20Analysis.table

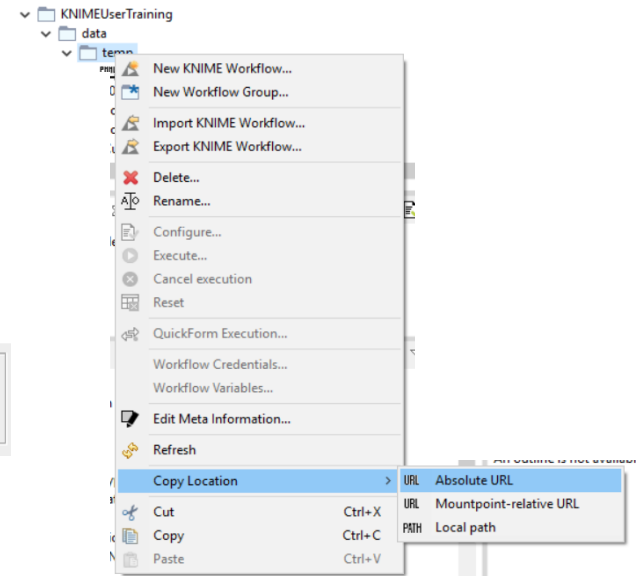
Browse...

Local path

Input location

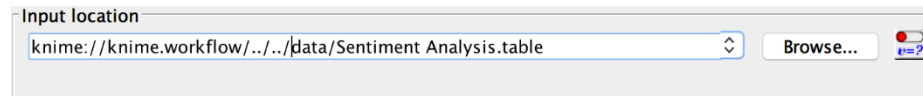
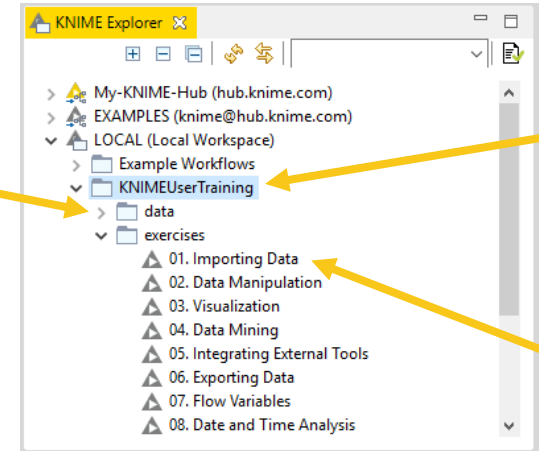
/Users/rb/knime-workspace/KNIMEUserTraining/data/Sentiment Analysis.table

Browse...



Workflow-Relative File Paths

- Best choice if workflows are to be shared
- Requires matching folder structure within workflow group
 - Independent of environment outside of workflow group
- Example: Path to „Sentiment Analysis.table“
 - Local path:
 - C:\Users\rb\knime-workspace\KNIMEUserTraining\data\Sentiment Analysis.table
 - Workflow relative:



YouTube KNIME TV Channel: <https://youtu.be/U9sP4g4yGwY>

New Node: Excel Reader (XLS)

- Reads .xls and .xlsx file from Microsoft Excel
- Supports reading from multiple sheets

Excel Reader (XLS)



**Read Excel Sheet
Names (XLS)**



Excel Reader Configuration

The screenshot shows the 'Excel Reader (XLS)' configuration dialog. It has four tabs: 'XLS Reader Settings' (active), 'Flow Variables', 'Job Manager Selection', and 'Memory Policy'. The 'Select file to read:' section has 'File' selected. The 'Read from:' dropdown is set to 'Mountpoint' and 'LOCAL'. The 'File:' field contains '/KNIMEUserTraining/data/Product Data2.xls'. A callout 'File path' points to this field. The 'Adjust Settings:' section includes 'Select the sheet to read:' set to '<first sheet with ...', 'Column Names:' with 'Table contains column names in row number: 1' checked, and 'Row IDs:' with 'Generate RowIDs (index incrementing, starting with "Row0")' selected. A callout 'Sheet specific settings' points to this section. The 'Select the columns and rows to read:' section has 'Read entire data sheet, or ...' checked. The 'On evaluation error:' section has 'Insert an error pattern: #XL_EVAL_ERROR#' selected. The 'More Options:' section has 'Skip empty columns', 'Skip hidden columns', and 'Skip empty rows' checked. The 'Preview' tab is active, showing a table with columns 'Row ID', 'Custo...', and 'Products'. A callout 'Preview' points to this table. The table data is as follows:

Row ID	Custo...	Products
Row0	11000	Private Investment
Row1	11001	Private Investment
Row2	11002	Private Investment
Row3	11003	Private Investment
Row4	11004	Private Investment

Filenames and the knime:// Protocol

Absolute URL

Read from: Mountpoint LOCAL

File: /KNIMEUserTraining/data/Product Data2.xls Browse...

Local Path

Read from: Local File System

File: /Users/kathrinmelcher/knime-workspace/KNIMEUserTraining/data/Product Data2.xls Browse...

Custom URL

Read from: Custom URL

URL: knime://knime.workflow/../../data/Product%20Data2.xls Browse...

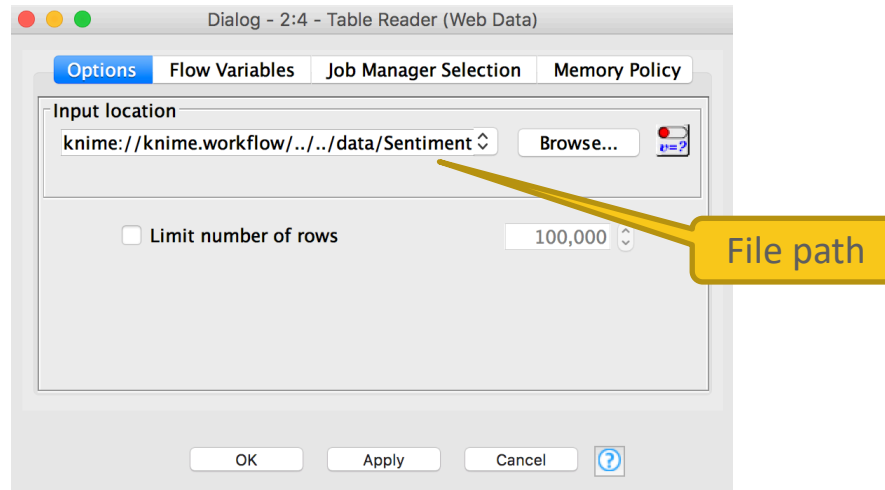
New Node: Table Reader

- Reads tables from the native KNIME Format.
- Maximum performance, minimum configuration

Table Reader



Web Data

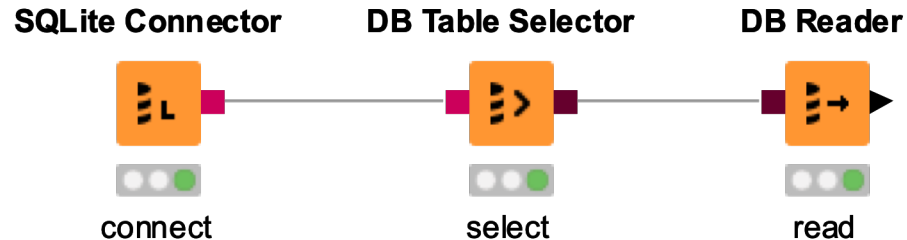


YouTube KNIME TV channel video:

<https://youtu.be/tid1qi2HAOo>

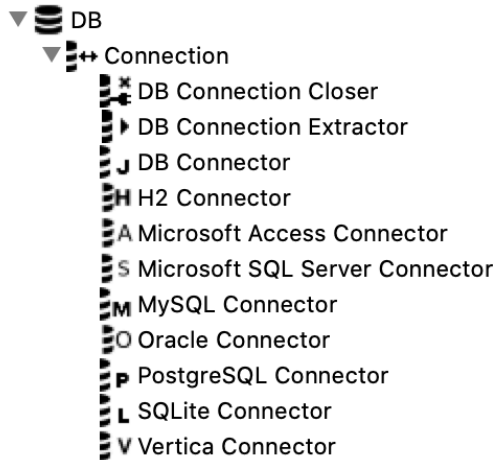
Database Connectivity

- Read data from any JDBC enabled database
- Write your own SQL or model it using dedicated nodes

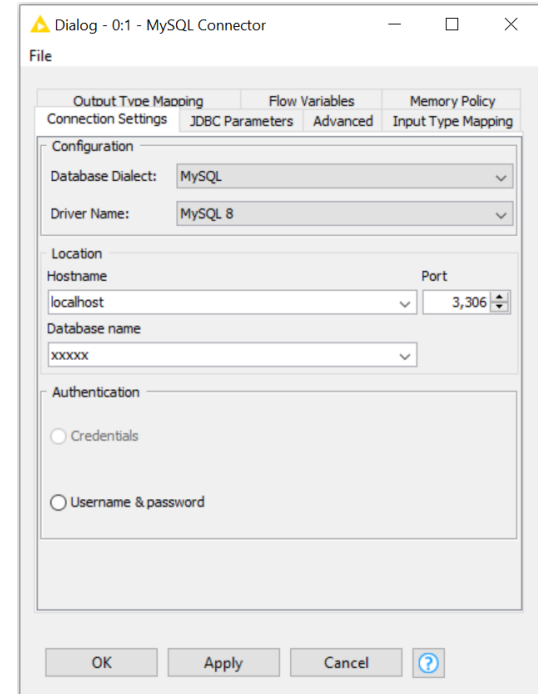


New Nodes: Database Connectors

- Native: Postgres, MySQL, MS SQL Server, SQLite
- DB Connector (e.g. DB2, HANA).
- Big Data: HIVE and Impala



MySQL Connector



Other Useful Data Sources

- PMML Reader – reads standard predictive models
- XML Reader with XPATH support
- Python/R Source nodes
- Tika Parser – extracts textual data from 200+ file types
- REST Web Services, and many more

XML Reader



CSV Reader



PMML Reader



R Source (Table)



Importing Data Exercise

Start with exercise: *Importing Data*

Read the following files

- *Sentiment Analysis.table*
- *Sentiment Rating.csv*
- *Product Data2.xls*

Optional: Read the *web_activity* table from the database *WebActivity.sqlite*

(hint: drag and drop the files from the KNIME Explorer panel to get started)

You can download the training workflows from the KNIME Hub:

<https://hub.knime.com/knime/space/Education/01%20KNIME%20User%20Training/>

Table Reader



Sentiment Evaluation
from KNIME

File Reader



Sentiment Rating
+
Sentiment Description

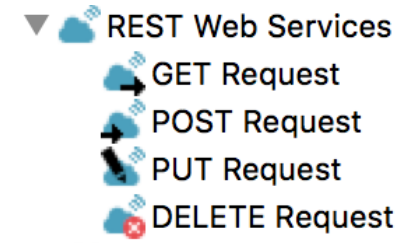
Excel Reader (XLS)



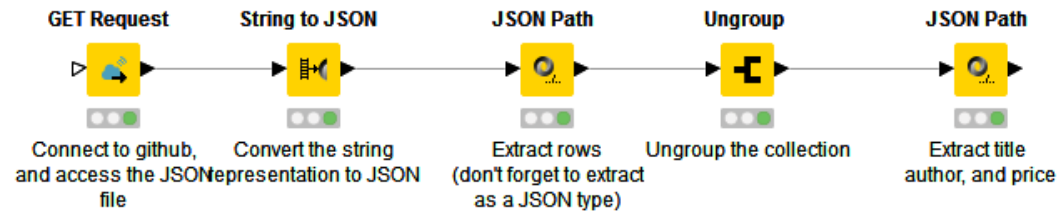
Products
<->
Customer

RESTful Web Services

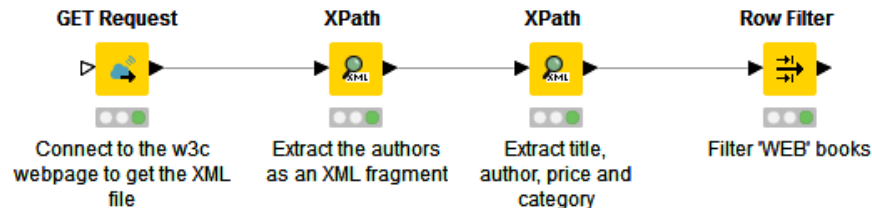
- Use KNIME nodes to interact with RESTful web services
- Send requests using standard HTTP methods



JSON Response:




XML Response:



RESTful Web Services

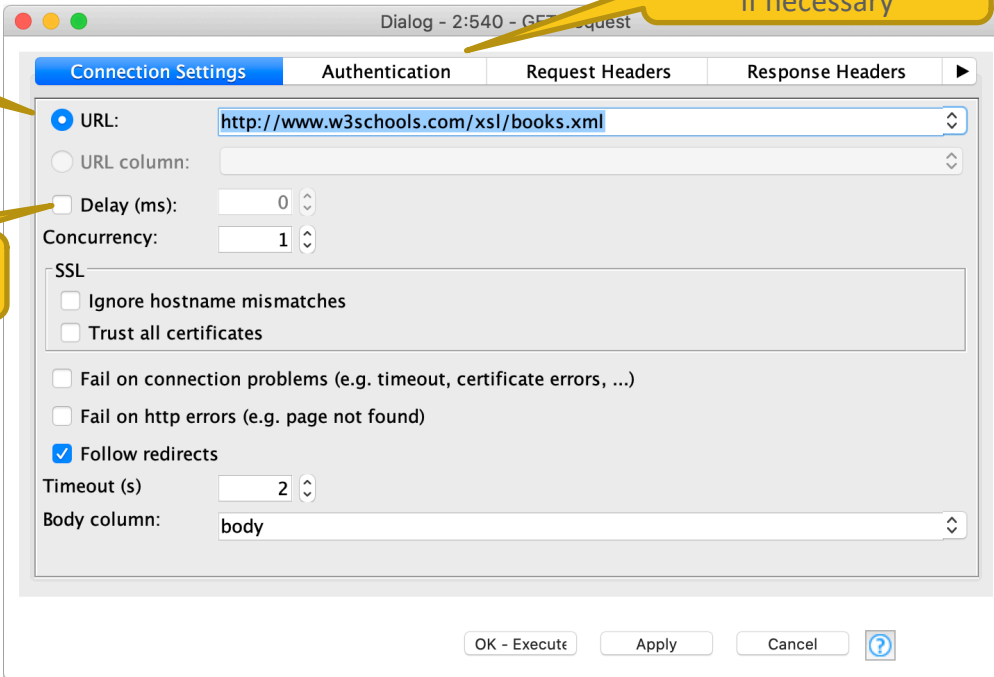
GET Request



Enter URL, or use from column

Add delay between individual requests

Provide authentication if necessary



Dialog - 2:540 - GET Request

Connection Settings Authentication Request Headers Response Headers

☒ URL:

☐ URL column:

☐ Delay (ms):

Concurrency:

SSL

☐ Ignore hostname mismatches

☐ Trust all certificates

☐ Fail on connection problems (e.g. timeout, certificate errors, ...)

☐ Fail on http errors (e.g. page not found)

☒ Follow redirects

Timeout (s)

Body column:

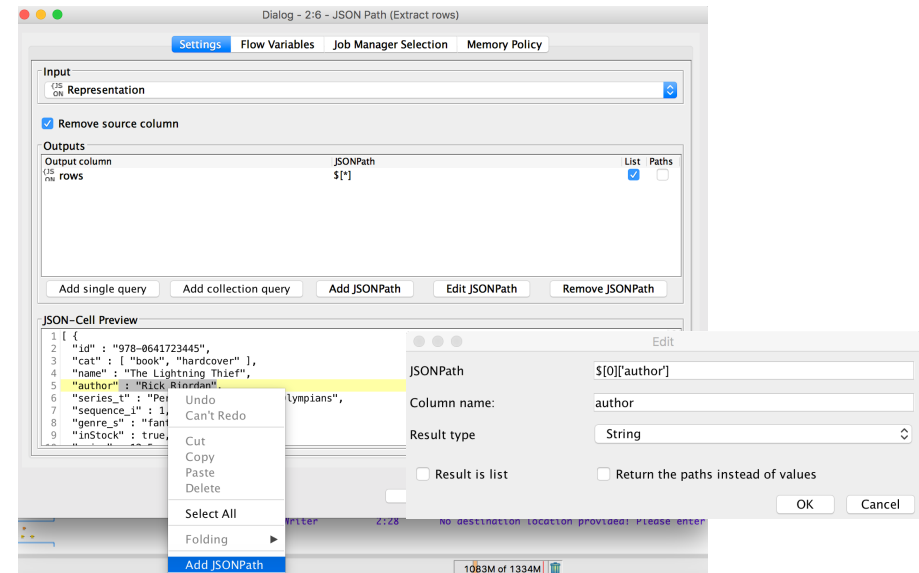
OK - Execute Apply Cancel ?

<https://www.knime.com/blog/a-restful-way-to-find-and-retrieve-data>

<https://www.knime.com/blog/OSM-meets-CSV-file-and-Google-API>

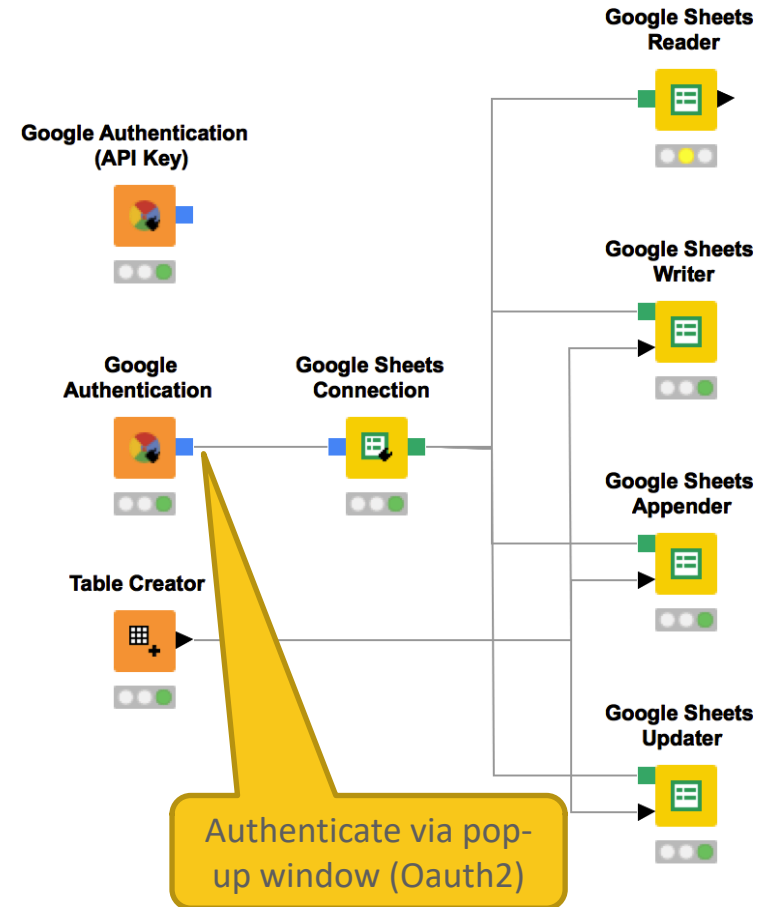
JSON Reader and JSON Path nodes

- Use the JSON Reader (or GET Request) node to get a JSON cell
- Use the JSON Path node to query the JSON file and extract parameters
- Editor window simplifies construction of JSON queries by auto-generating them (click on properties)



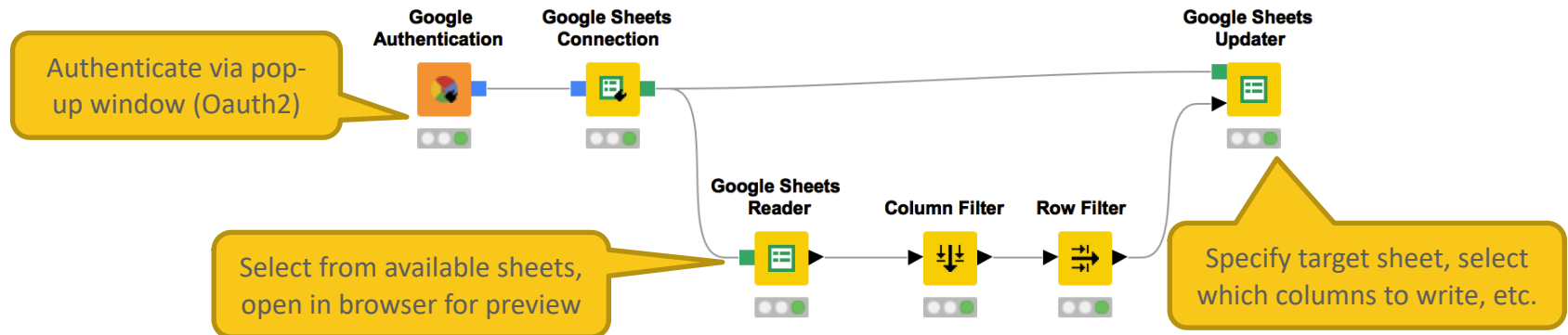
Google Sheets

- Access your data stored in Google Services
 - Read data from Google Sheets
 - Write data to new sheets
 - Modify existing sheets
- Makes collaboration and sharing of data easy
 - (especially vs. sending Excel sheets via email...)



Google Sheets

- Select from available sheets on Google Drive
- Transform data in KNIME, or enrich with new data
- Create new sheet or update existing sheets
 - Allows to read from / write to specific range of sheet (e.g. A1:G10)

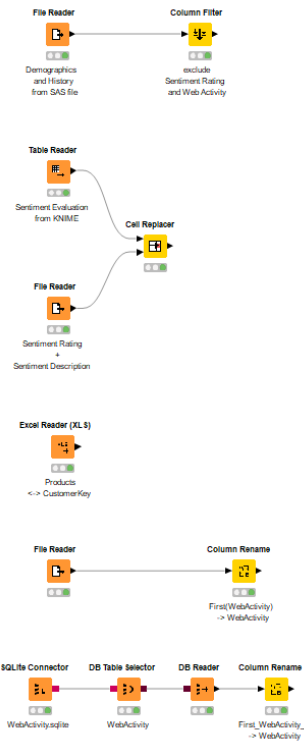


Today's Example

Final Workflow from the KNIME User Training

... and putting all those parts together, you get this final workflow.

Data Reading



Visualization

Data Manipulation and Aggregation

Training Predictive Models

Data Export and Reporting

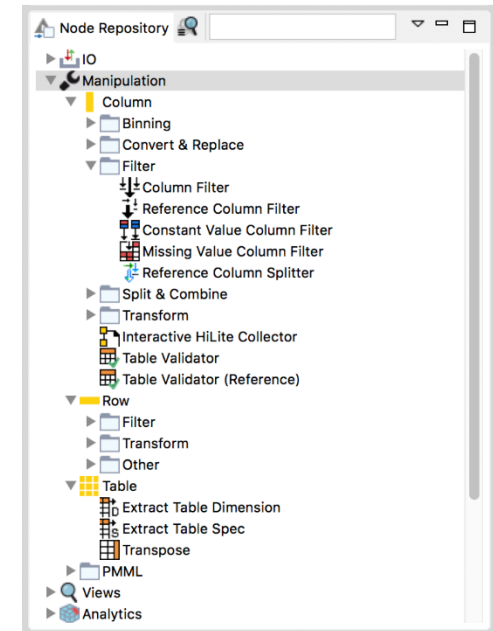
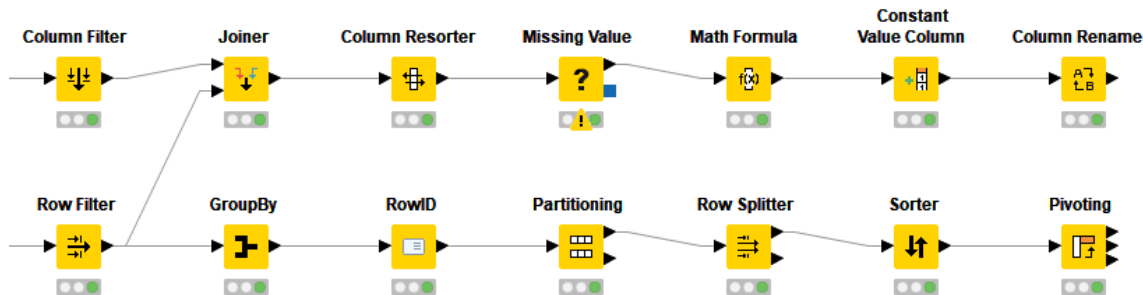
Data Manipulation

Clean, Join, Aggregate



Data Manipulation Nodes

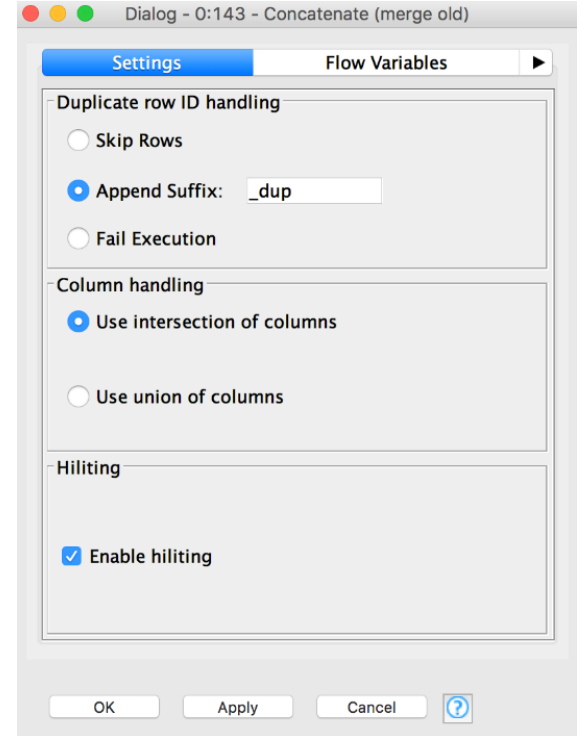
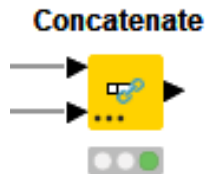
- Yellow color with a variety of input and output ports
- Apply a transformation to input data
- Many, many nodes!



New Node: Concatenate

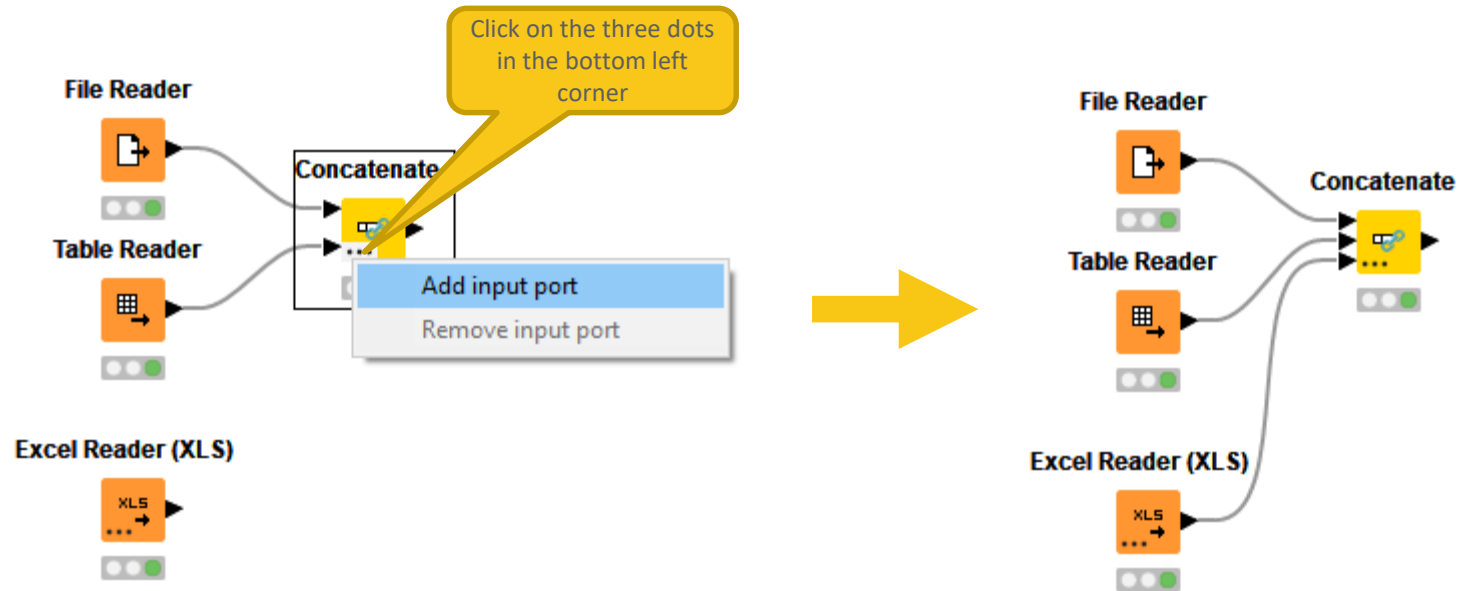
Combine rows from 2 or more tables with shared columns

- Handles duplicate row keys gracefully
- Take the union or intersection of columns



Dynamic Ports

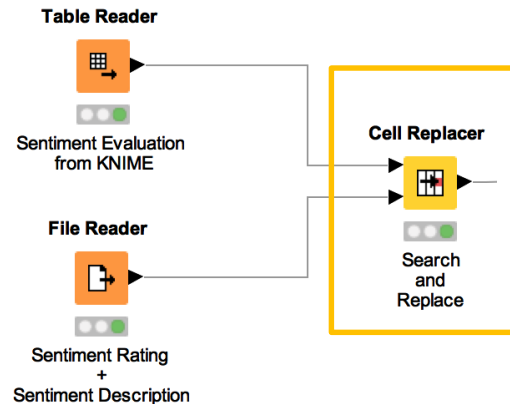
Add and remove node ports based on your needs, e.g. in order to concatenate three or more tables



New Node: Cell Replacer

Replaces the content of a column based on a lookup

- Top port references the table to be searched
- Bottom port holds the lookup table (search keys and replacement values)



The screenshot shows the 'Dialog - 0:129 - Cell Replacer (Search)' window. It has two tabs: 'Options' (selected) and 'Flow Variables'. The 'Options' tab contains the following settings:

- Input table:** Target column is set to 'Sentiment Analysis'.
- Dictionary table:** Input (Lookup) is set to 'Sentiment Analysis' and Output (Replacement) is set to 'SentimentRating'.
- Append/Replace Result Column:** The checkbox 'Append new column' is checked, and the column name is 'SentimentRating'.
- If no element matches use:** The radio button 'Missing' is selected.

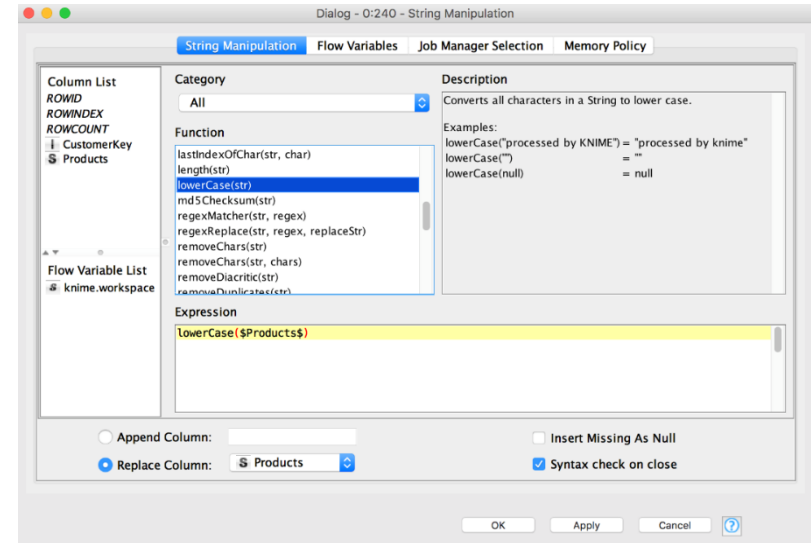
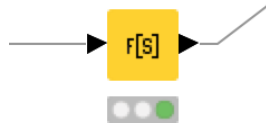
At the bottom, there are buttons for 'OK', 'Apply', 'Cancel', and a help icon.

New Node: String Manipulation

Create and edit values in String columns

- Clean up capitalization (eg. Lowercase)
- Replace strings
- Modify existing strings or create new columns

String Manipulation



Data Manipulation Exercise, Activity I

Start with exercise: *Data Manipulation, Activity I*

- Concatenate web activity data from the old and new systems
- Replace the written sentiment values with the numeric sentiment scores
- Make sure that all product names in the product data spreadsheet are written in lower case letters

Joining Columns of Data

Left Table

CustomerKey	OrderDate	OrderID
22	2019-09-23	#23444
24	2019-09-30	#23457
15	2019-10-07	#28985
10	2091-10-13	#29999

Join by CustomerKey

Inner Join

Right Table

CustomerKey	DoB	City	Gender
17	1974-02-23	Berlin	F
65	2001-05-25	Stuttgart	F
35	1988-08-05	Cologne	M
15	1983-07-20	Hamburg	M
10	1993-01-13	Berlin	M

Left Outer Join

CustomerKey	OrderDate	OrderID	DoB	City	Gender
15	2019-10-07	#28985	1983-07-20	Hamburg	M
10	2091-10-13	#29999	1993-01-13	Berlin	M

Right Outer Join

CustomerKey	OrderDate	OrderID	DoB	City	Gender
22	2019-09-23	#23444	?	?	?
24	2019-09-30	#23457	?	?	?
15	2019-10-07	#28985	1983-07-20	Hamburg	M
10	2091-10-13	#29999	1993-01-13	Berlin	M

CustomerKey	OrderDate	OrderID	DoB	City	Gender
17	?	?	1974-02-23	Berlin	F
65	?	?	2001-05-25	Stuttgart	F
35	?	?	1988-08-05	Cologne	M
15	2019-10-07	#28985	1983-07-20	Hamburg	M
10	2091-10-13	#29999	1993-01-13	Berlin	M

Joining Columns of Data

Left Table

CustomerKey	OrderDate	OrderID
22	2019-09-23	#23444
24	2019-09-30	#23457
15	2019-10-07	#28985
10	2091-10-13	#29999

Join by CustomerKey

Right Table

CustomerKey	DoB	City	Gender
17	1974-02-23	Berlin	F
65	2001-05-25	Stuttgart	F
35	1988-08-05	Cologne	M
15	1983-07-20	Hamburg	M
10	1993-01-13	Berlin	M

Full Outer Join

Missing values in the left table

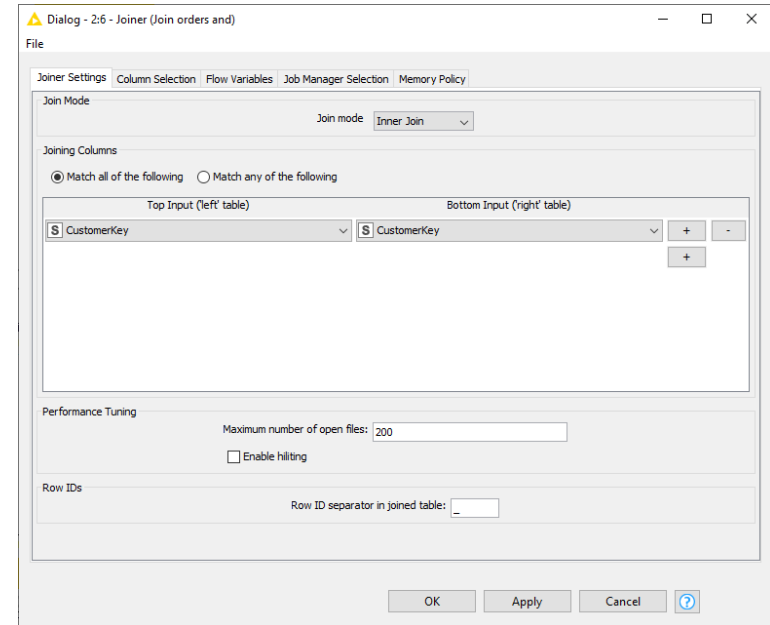
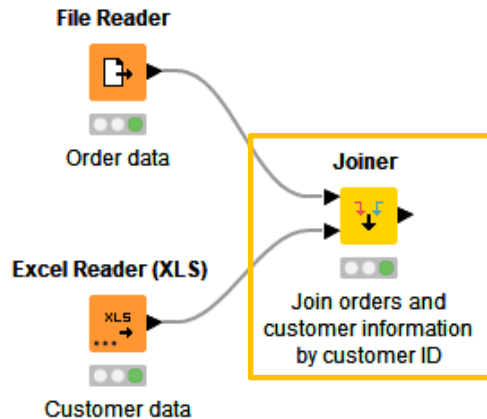
CustomerKey	OrderDate	OrderID	DoB	City	Gender
17	?	?	1974-02-23	Berlin	F
65	?	?	2001-05-25	Stuttgart	F
35	?	?	1988-08-05	Cologne	M
15	2019-10-07	#28985	1983-07-20	Hamburg	M
10	2091-10-13	#29999	1993-01-13	Berlin	M
22	2019-09-23	#23444	?	?	?
24	2019-09-30	#23457	?	?	?

Missing values in the right table

New Node: Joiner

Combines columns from 2 different tables

- Top port contains “Left” data table
- Bottom port contains “Right” data table



Joiner Configuration – Linking Rows

Dialog - 2:6 - Joiner (Join orders and)

File

Joiner Settings | Column Selection | Flow Variables | Job Manager Selection | Memory Policy

Join Mode

Join mode: Inner Join

Joining Columns

☒ Match all of the following ☐ Match any of the following

Top Input ('left' table)	Bottom Input ('right' table)
S CustomerKey	S CustomerKey

Performance Tuning

Maximum number of open files: 200

☐ Enable hlling

Row IDs

Row ID separator in joined table: _

OK Apply Cancel ?

Joiner mode

Values to join on.
Multiple joining columns
are allowed

Joiner Configuration – Column Selection

Dialog - 2:6 - Joiner (Join orders and)

File

Joiner Settings **Column Selection** Flow Variables Job Manager Selection Memory Policy

Top Input (left table)

Exclude

Filter

No columns in this list

Include

Filter

CustomerKey
OrderDate
OrderID

☒ Always include all columns

Bottom Input (right table)

Exclude

Filter

City
Gender

Include

Filter

CustomerKey
DoB

☐ Always include all columns

Duplicate Column Handling

☐ Filter duplicates
☐ Don't execute
☒ Append suffix (automatic)
☐ Append custom suffix: (*)

Joining Columns Handling

☐ Remove joining columns from top input (left table)
☒ Remove joining columns from bottom input (right table)

OK Apply Cancel ?

Columns from left table to output table

Columns from right table to output table

Data Aggregation

Product ID	Category	# Ordered Items
P 1	Clothing	2
P 2	Home	3
P 3	Clothing	1
P 4	Clothing	5
P 5	Electronics	7
P 6	Electronics	5



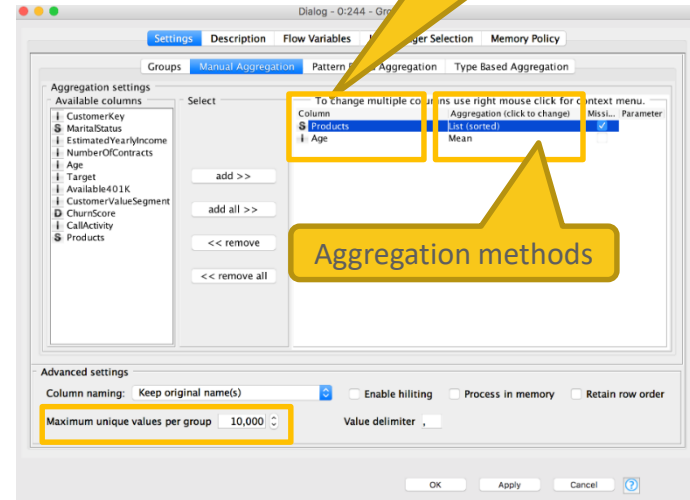
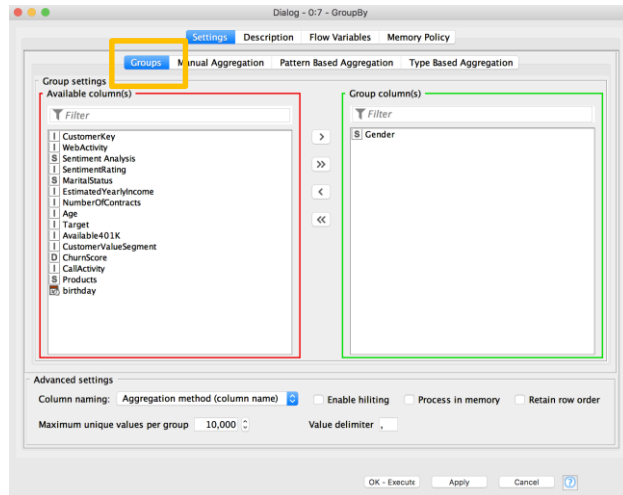
Group	Sum(# Ordered Items)
Clothing	8
Home	3
Electronics	12

Aggregated on Category (group) by Sum (aggregation method)

New Node: GroupBy

Aggregate rows to summarize data

- First tab provides grouping options
- Second tab provides control over aggregation details



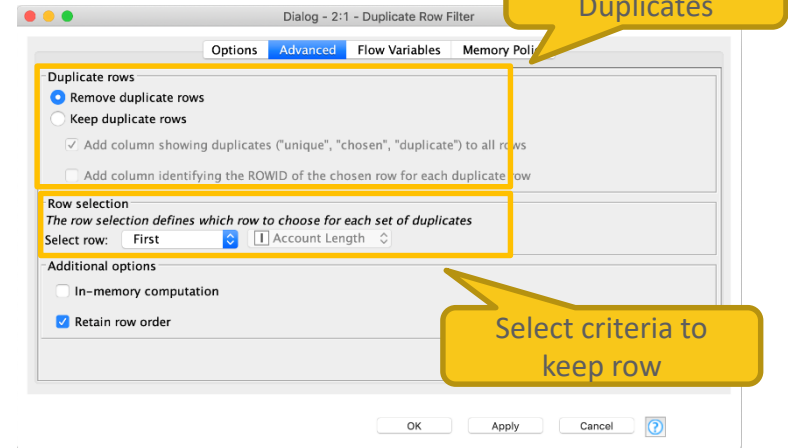
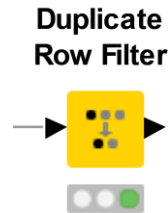
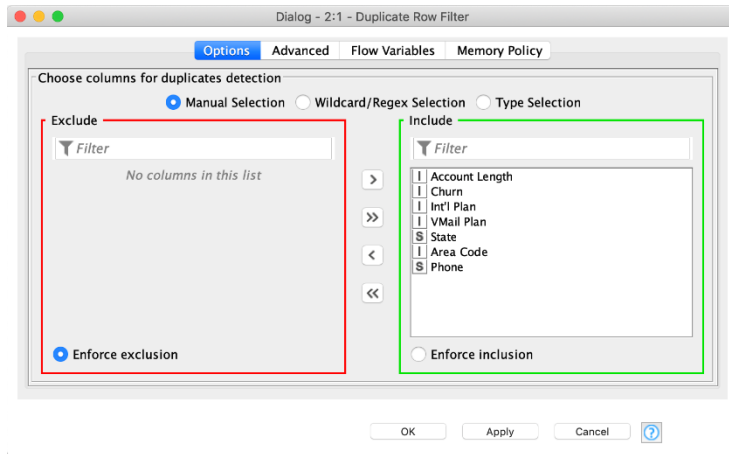
YouTube KNIME TV video:

<https://youtu.be/bDwF-TOMtWw>

New Node: Duplicate Row Filter

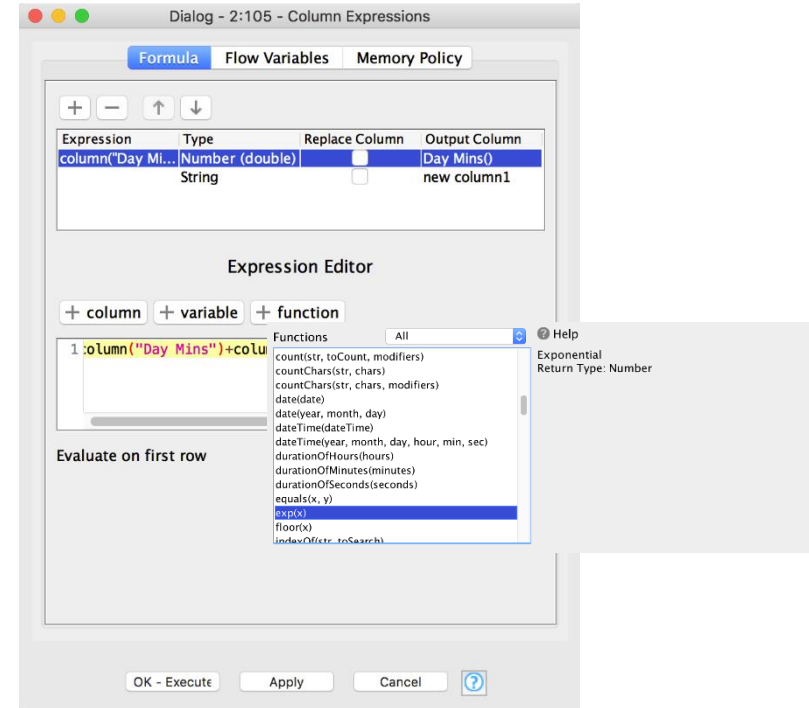
Detect duplicate row and apply a selected treatment

- First tab provides the option to select columns
- Second tab provides options for treating duplicated values



New Node: Column Expression

- Append or modify an arbitrary number of columns using expressions
- Many different functions are available
- No restriction on number of lines per expression allow to write complex expressions
- Part of the KNIME Labs extension

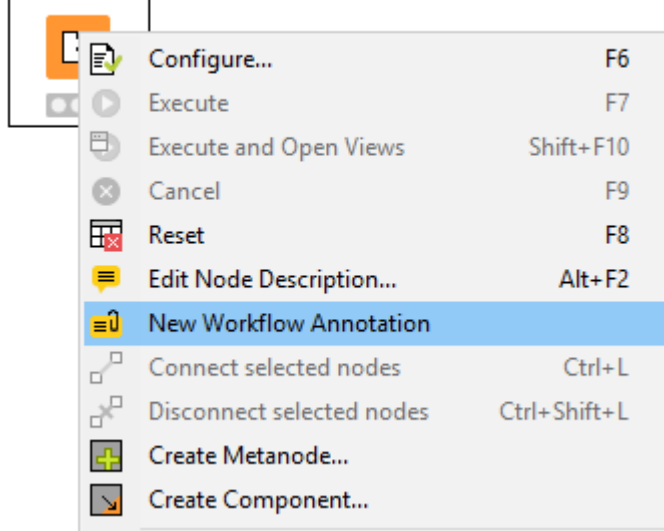


Workflow Organization and Documentation



Comments & Annotations

File Reader



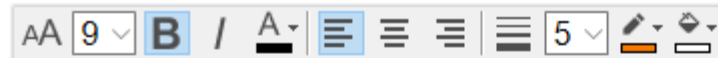
Double-click to write
Use the panel to
change properties

File Reader

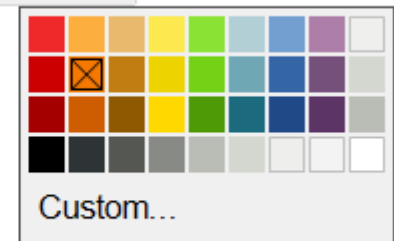


This node reads
the contract data
in **folder**
two levels up
from the folder
where the workflow
is currently executing

This is an example workflow annotation.
Here I can describe the **task** of a group of
nodes



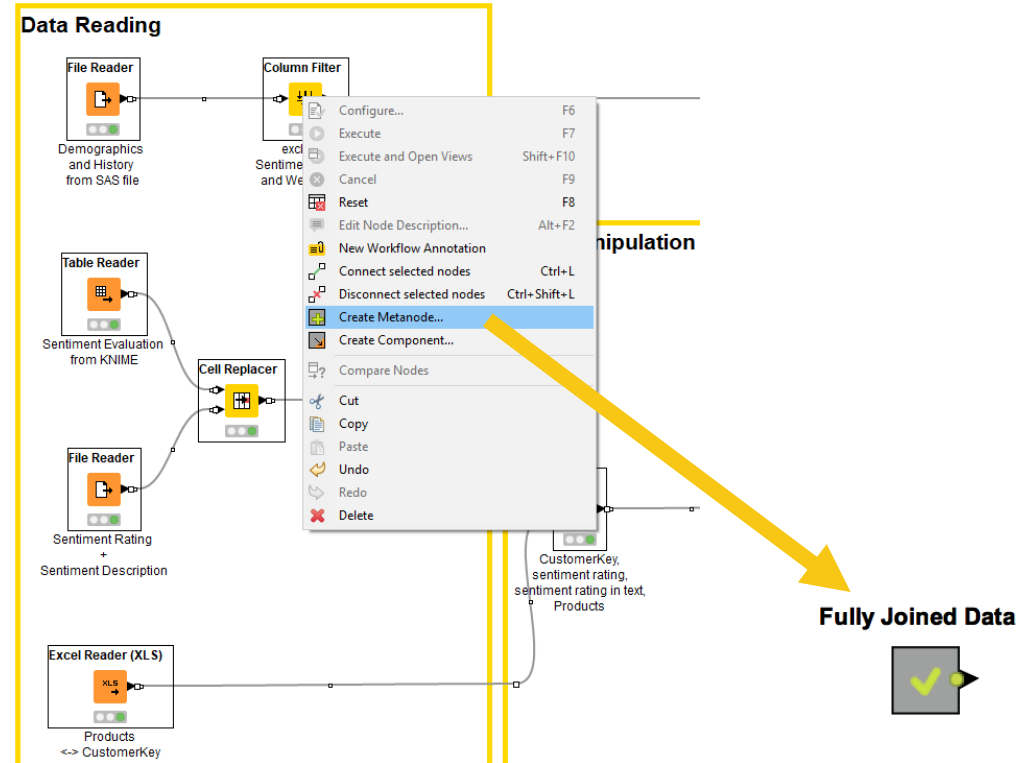
This is an example workflow annotation.
Here I can describe the **task** of a group of
nodes



YouTube KNIME TV Channel:
https://youtu.be/AHURYB_O8sA

Workflow Organisation – Good Practices

- Workflow annotations
- Node labels
- Metanodes
 - Right click -> Create Metanode...
 - Organize workflow by task
 - Hide complexity & improve readability



Workflow Organisation – Components

- Component encapsulates a reusable functionality as a KNIME workflow
- Components can be configured as any KNIME nodes
- Access and share components on the KNIME Hub

The screenshot shows the KNIME Hub search results for 'column filter'. The search bar at the top contains 'column filter'. Below the search bar, it says '36 results'. The 'Components' tab is selected and circled in yellow. Two components are listed: 'Interactive Column Filter' and 'Column Filter (by Index)'. A yellow speech bubble points to the 'Column Filter (by Index)' component, stating: 'Drag and drop from the KNIME Hub to your workflow'. Below the speech bubble, a small icon of the 'Column Filter (by Index)' component is shown. To the right, a workflow diagram for the 'Column Filter (by Index)' component is displayed. The workflow starts with a 'Component Input' node, followed by a 'String Configuration' node, a 'Variable to Table Row' node, and a 'Cell Splitter' node. The output of the 'Cell Splitter' is connected to an 'Error treatment' node. The output of the 'Error treatment' node is connected to a 'Select column subset' node. The output of the 'Select column subset' node is connected to a 'Reference Column Filter' node, which finally leads to a 'Component Output' node. A text box at the bottom of the workflow diagram provides input and extension details.

Drag and drop from the KNIME Hub to your workflow

Column Filter (by Index)

Get a table with column indices

String Configuration

Variable to Table Row

Cell Splitter

Error treatment

Handle situations, when something went wrong

Component Input

Select column subset

Reference Column Filter

Component Output

Input is a string of comma-separated column indices.
Negative numbers mean counting from the back, e.g. -1 is the last column.
Zeros is not a valid input!
Examples of the configuration:
1,3,5: 1st, 3rd and 5th columns
1,-1: 1st and last columns
The component requires the following extensions:
- KNIME Expressions (<https://hub.knime.com/knime/extensions/org.knime.features.expressions/latest>)

KNIME WorkflowDiff

- Automates identification and comparison of nodes in a workflow, metanodes, and two different workflows
- Identifies insertions, deletions, substitutions, and parameter changes

Name	Type	Value
Node Settings		
column-filter	sub-config	
filter-type	string	STANDARD
included_names	sub-config	
array-size	int	3
0	string	petal length
1	string	petal width
2	string	class
excluded_names	sub-config	
enforce_option	string	EnforceExclusion
name_pattern	sub-config	
datatype	sub-config	
System Node Settings		

Name	Type	Value
Node Settings		
Document Column Internals	sub-config	
Document Column	string	Preprocessed Document
Preprocess Unmodifiable_Inte	sub-config	
Preprocess Unmodifiable	boolean	false
Replace Document Internals	sub-config	
Replace Document	boolean	true
New Document Column Nam	sub-config	
New Document Column Nam	string	Preprocessed Document
Stemmer Name Internals	sub-config	
Stemmer Name	string	Porter
System Node Settings		

Data Manipulation Exercise, Activity II

Start with exercise *Data Manipulation, Activity II*

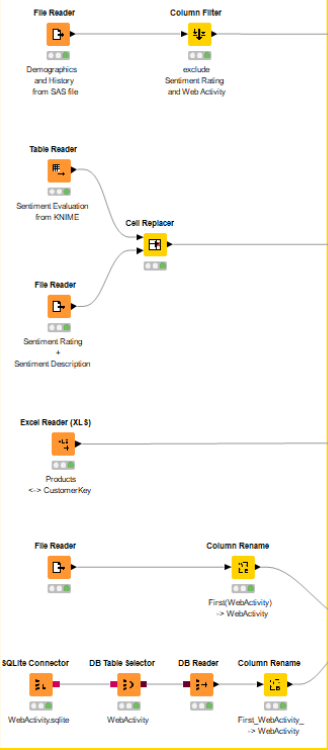
- Join all data into one table using a series of joiner nodes (use "Customer Key" as the joining column)
- Filter out duplicate rows
- Clean up and document your workflow using annotations, node labels, and metanodes

Today's Example

Final Workflow from the KNIME User Training

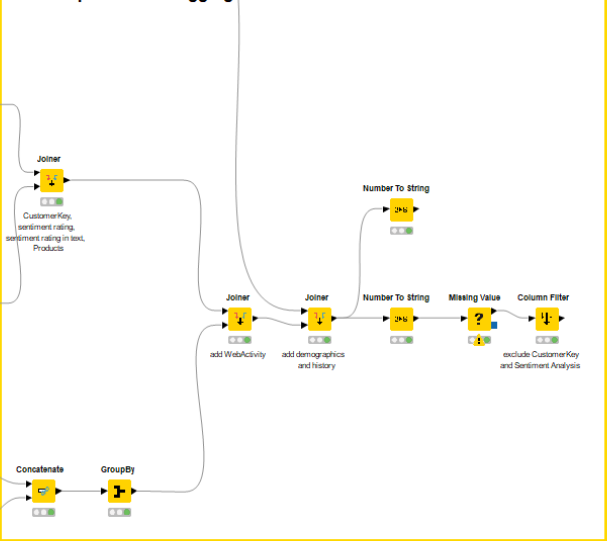
...and putting all those parts together, you get this final workflow.

Data Reading



Visualization

Data Manipulation and Aggregation



Training Predictive Models

Data Export and Reporting

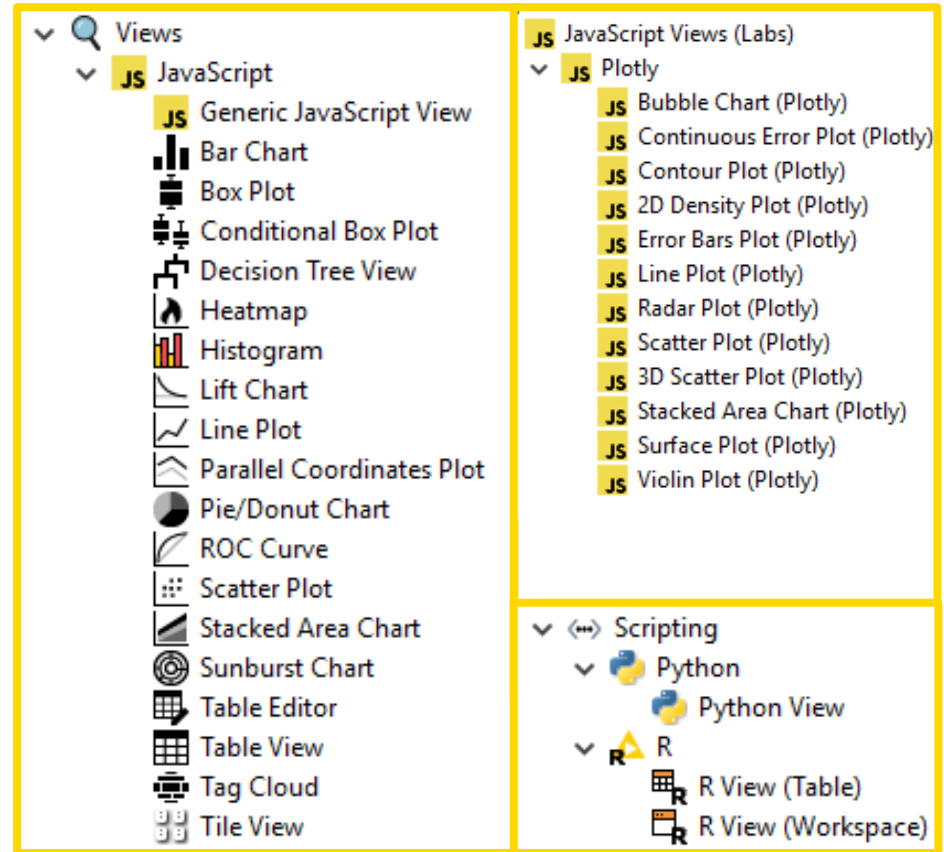
Data Visualization

Charts and Tables



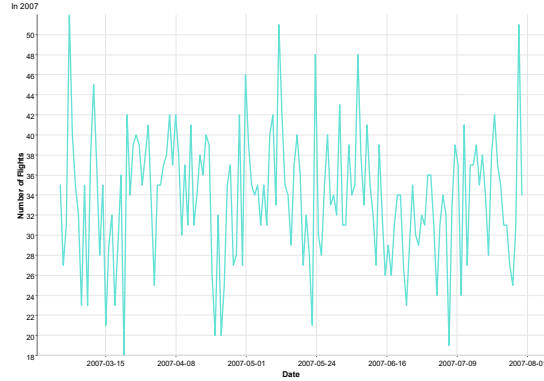
Data Visualization

- Large selection of easy to use visualization nodes
 - Web-based and interactive
 - Dedicated nodes, no scripting required
- Plotly nodes
 - Similar but integrated from an external library
- R and Python View nodes for highly customizable graphics
 - Require scripting

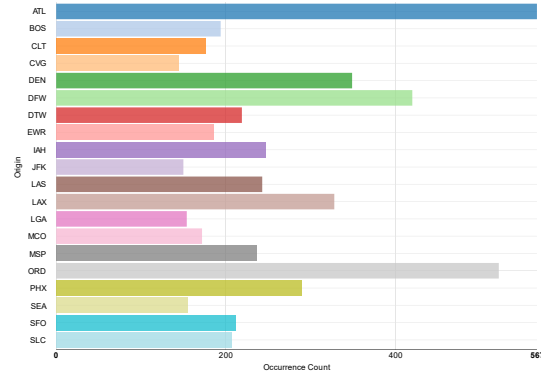


Visualizations using 1 Column

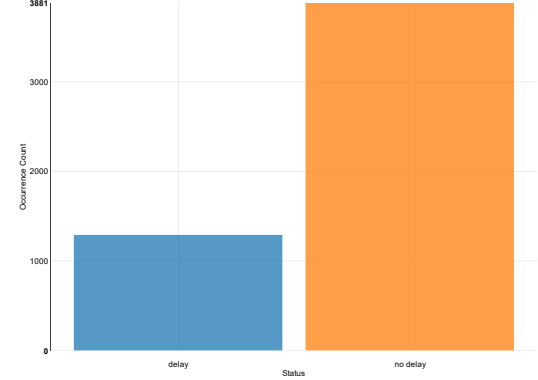
Number of Flights by Date



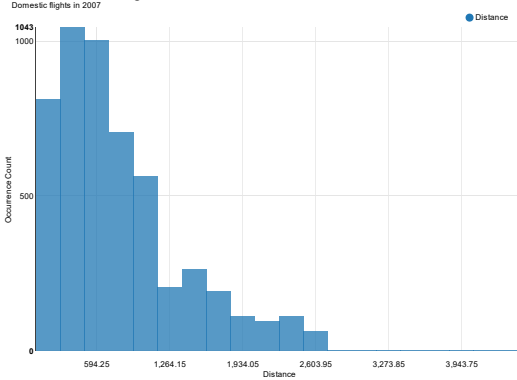
Departure Airports



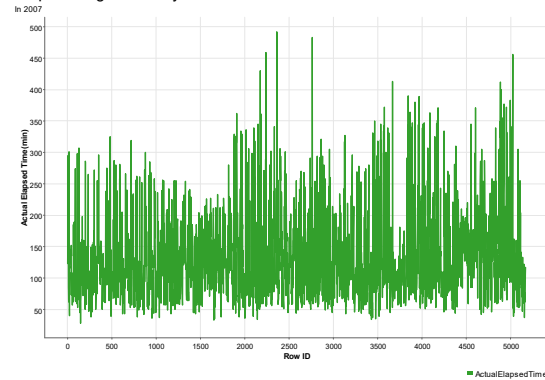
Distribution of Delayed and Non-Delayed Flights



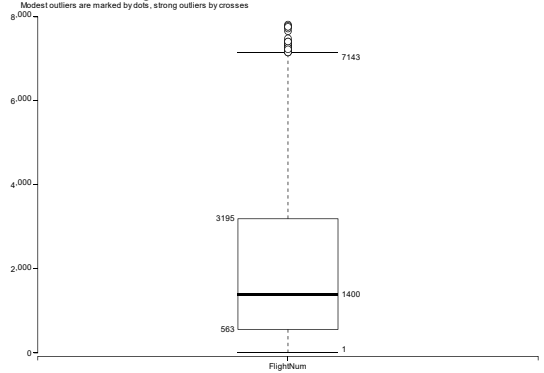
Distribution of Flight Distances



Elapsed Flight Time by Row ID

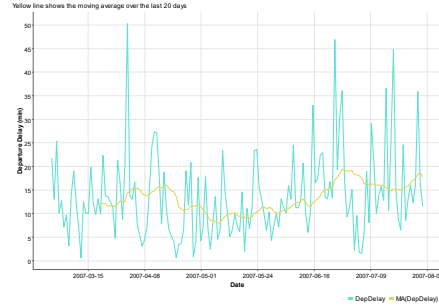


Distribution of Flight Numbers

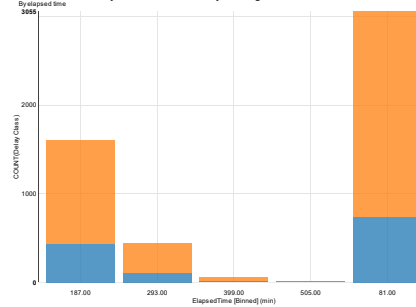


Visualizations using 2 Columns

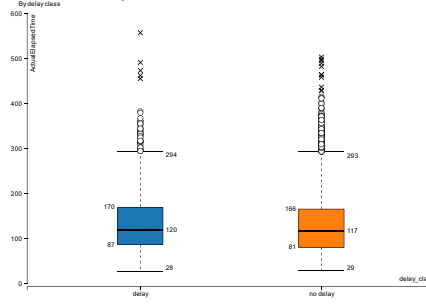
Delay Times by Date



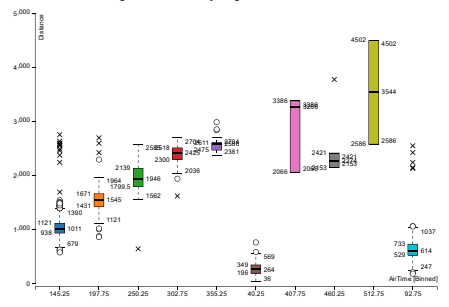
Number of Delayed and Non-Delayed Flights



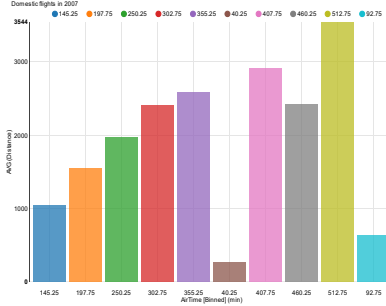
Distribution of Elapsed Time



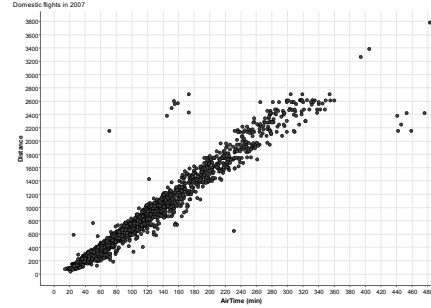
Distribution of Flight Distances by Flight Times



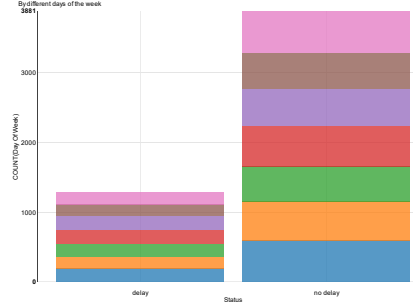
Average Distances by Flight Times



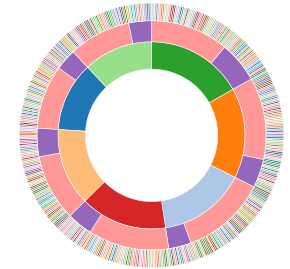
Correlation between Flight Time and Distance



Number of Delayed and Non-Delayed Flights

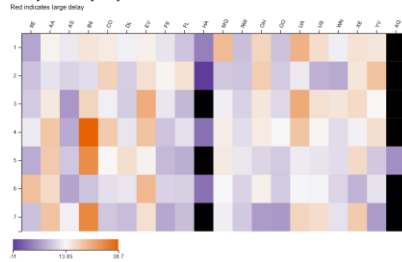


Flights by Day of the Week and Delay Status

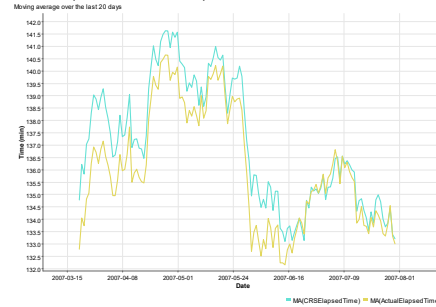


Visualizations using 3 Columns

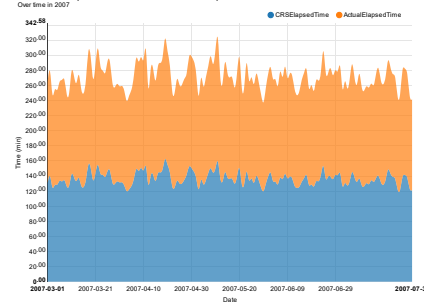
Delay Times by Day of the Week and Carrier



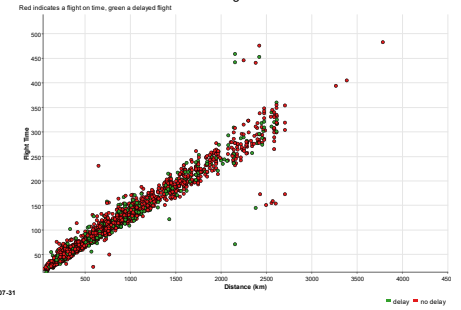
Actual Elapsed Time vs CRS Elapsed Time



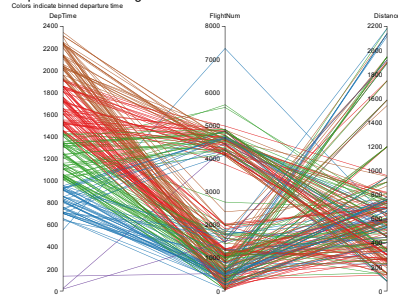
Actual Elapsed Time and CRS Elapsed Time



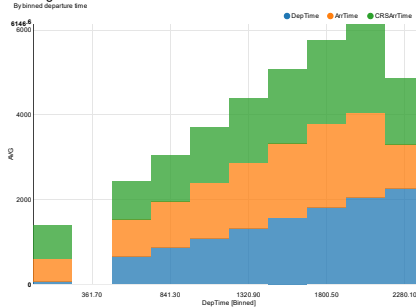
Correlation between Distance and Flight Time



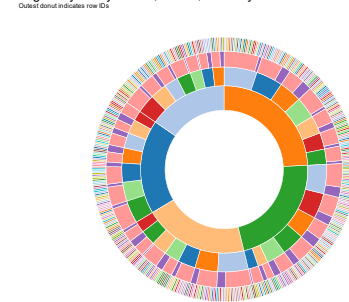
Departure Time vs Flight Number vs Distance



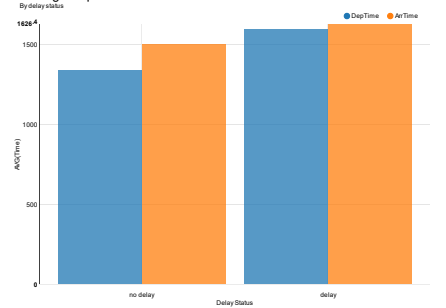
Average Arrival Time and CRS Arrival Time



Flights by Delay Status, Month, and Day of the Week

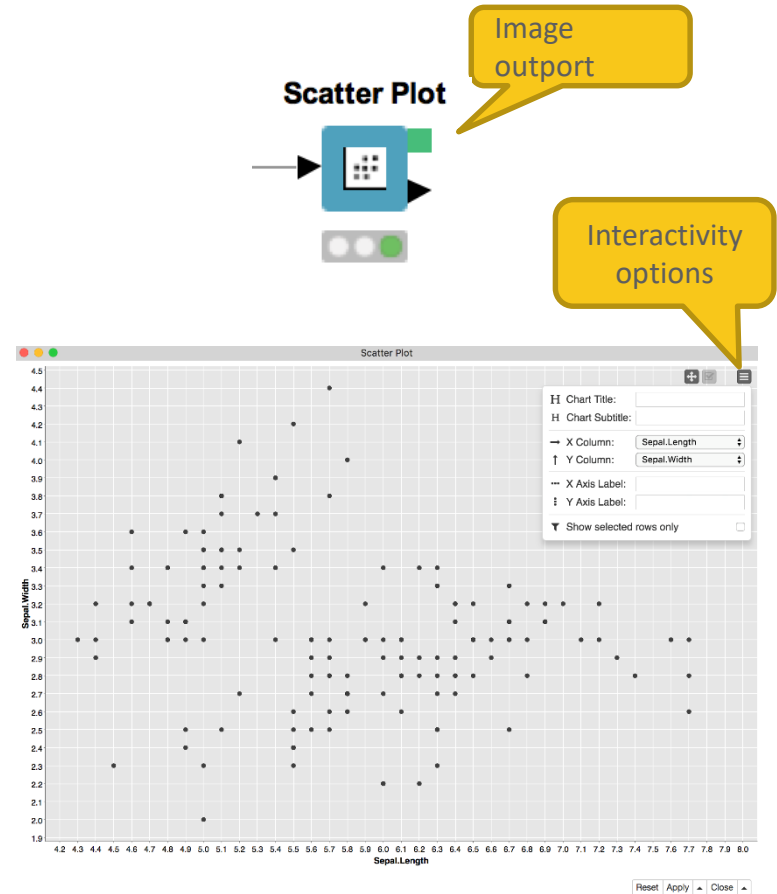


Average Departure and Arrival Times



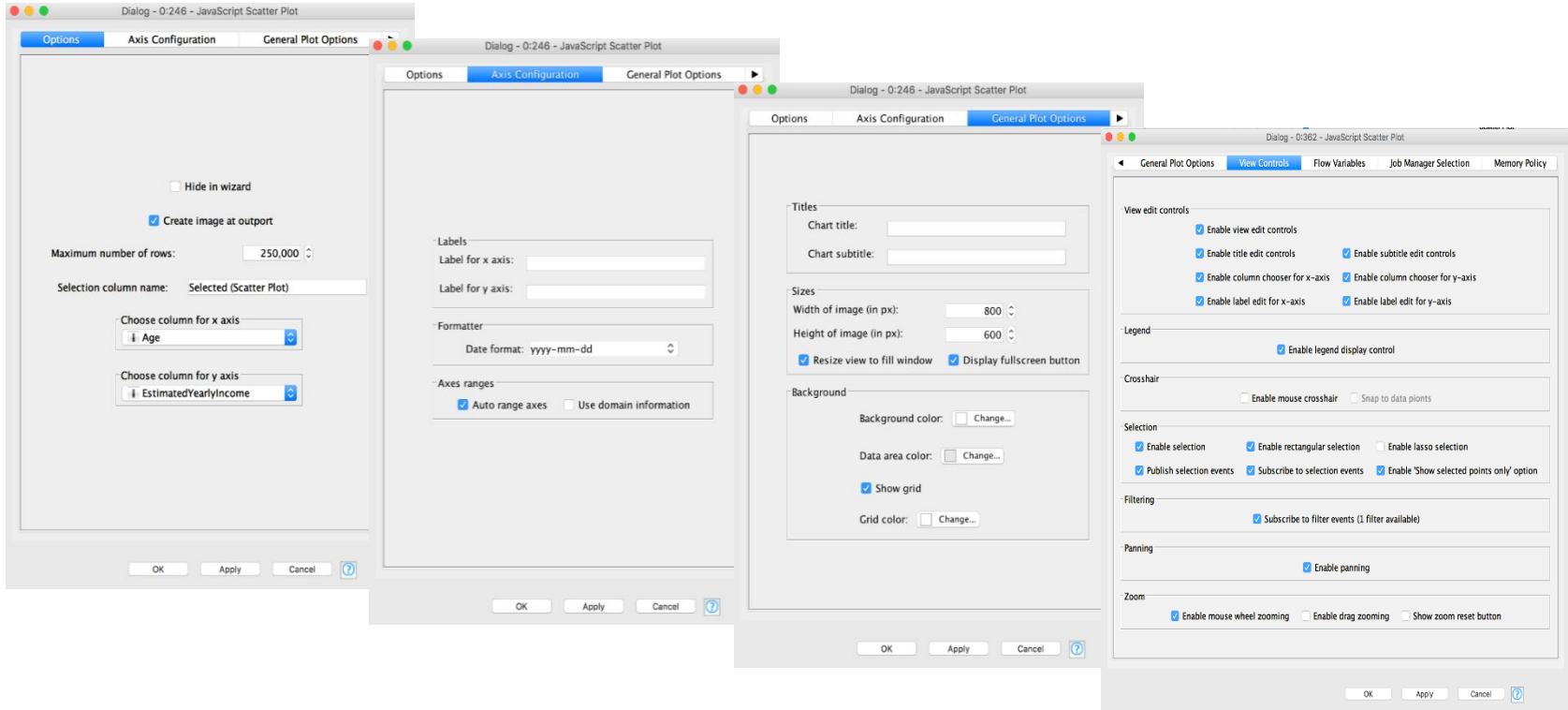
New Node: Scatter Plot

- Plots different columns on X and Y
- Displays data including color information
- Produces an interactive view and an image
- Select data points and publish selection to other views



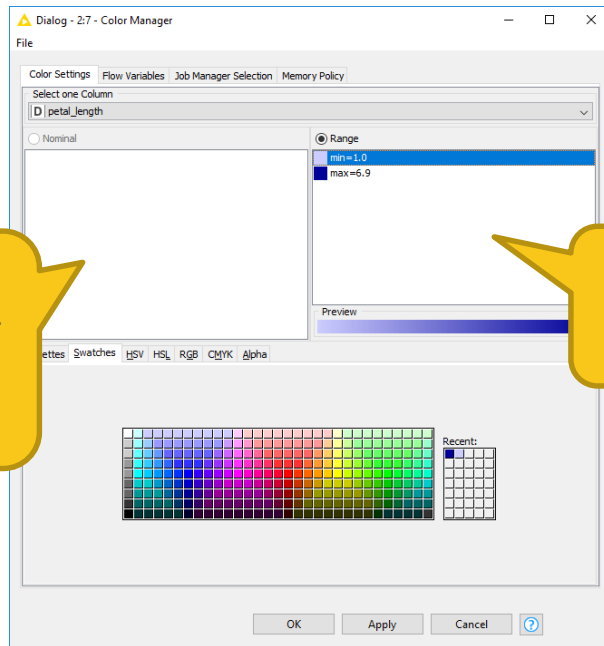
New Node: Scatter Plot

Four configuration tabs



New Node: Color Manager

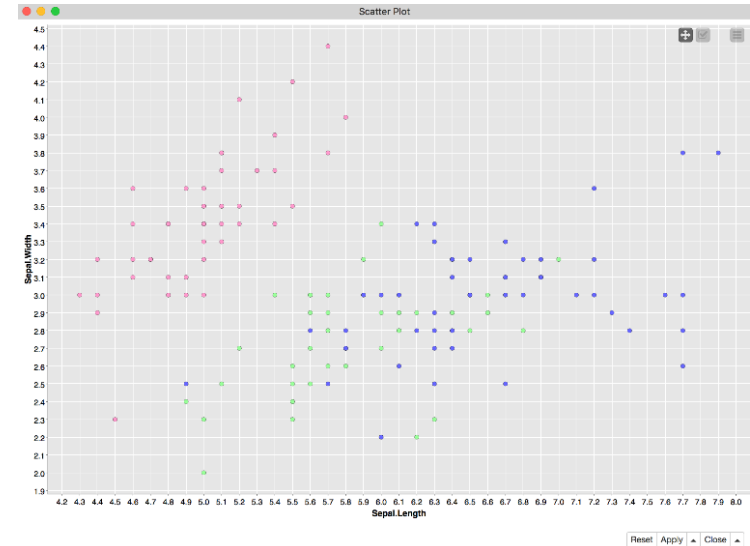
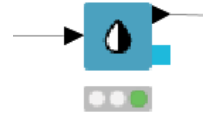
- Color by nominal or continuous values
- Sync colors between views using the color model port and Color Appender node



Discrete colors for nominal values

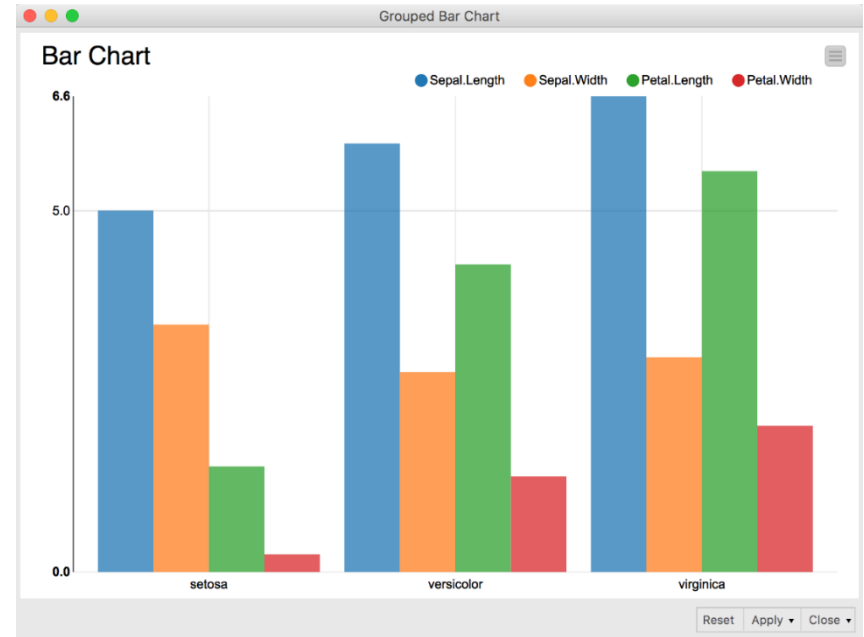
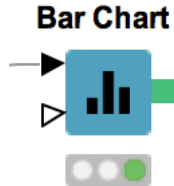
Color range for numerical values

Color Manager



New Node: Bar Chart

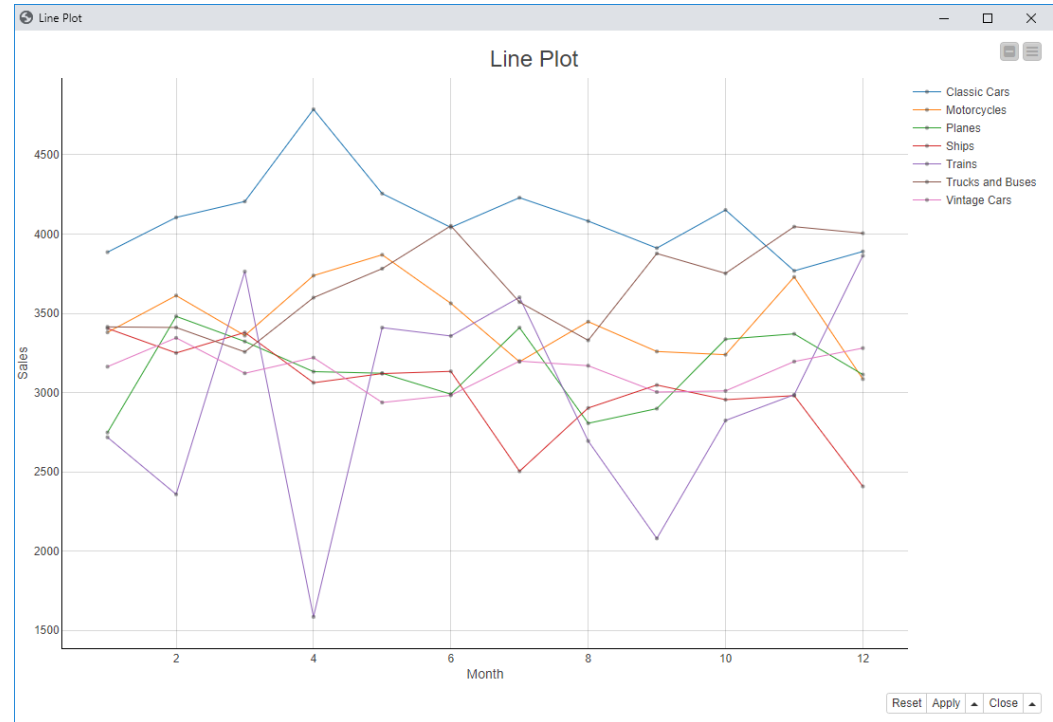
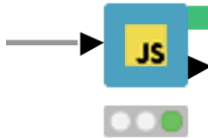
- Show numerical values across categories
- Vertical or horizontal bars
- Bars can be grouped or stacked



New Node: Line Plot

- Plot sequence of values, e.g. over time
- Useful to identify trends, also between groups

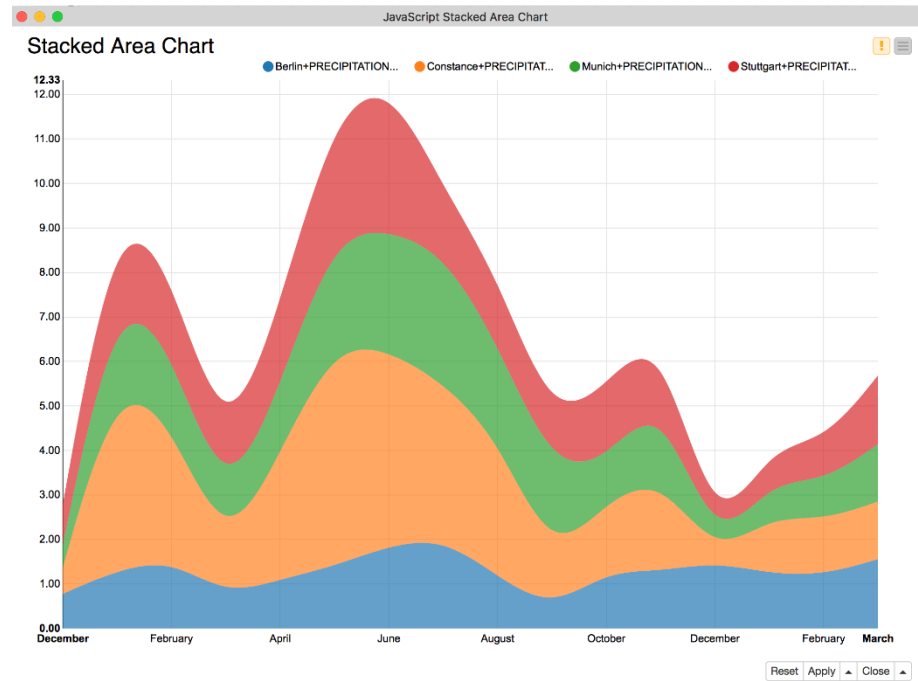
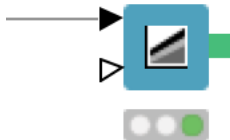
Line Plot (Plotly)



New Node: Stacked Area Chart

- Visualizes numerical values from multiple columns as stacked areas
- Great for plotting distributions over time

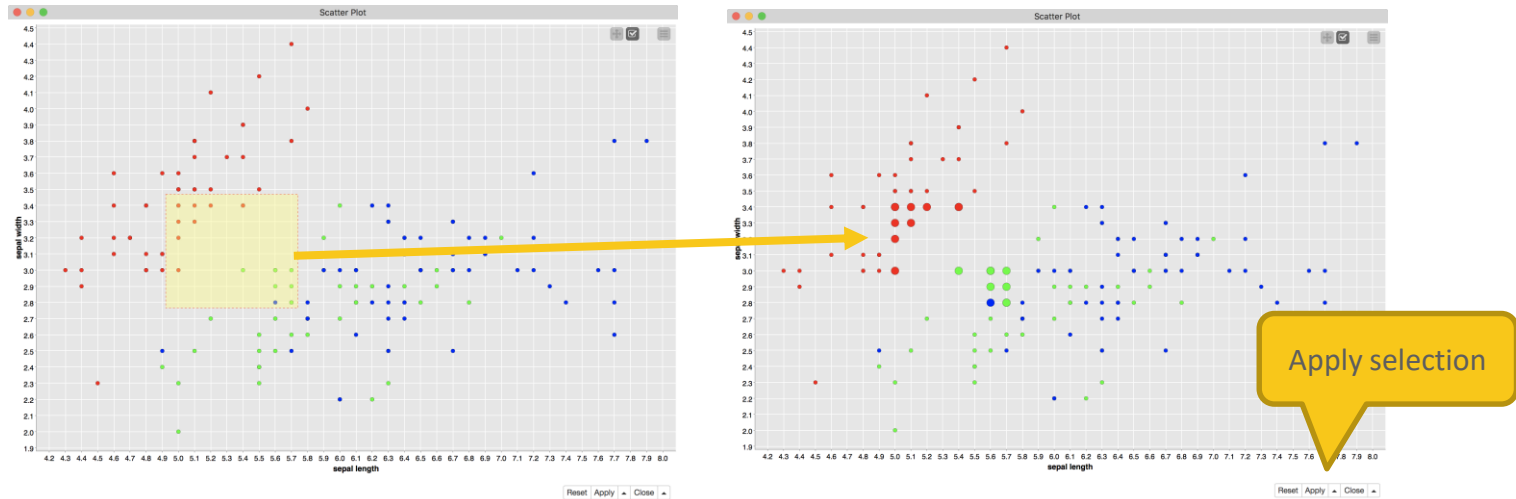
Stacked Area Chart



Selection & Filtering in JavaScript Views

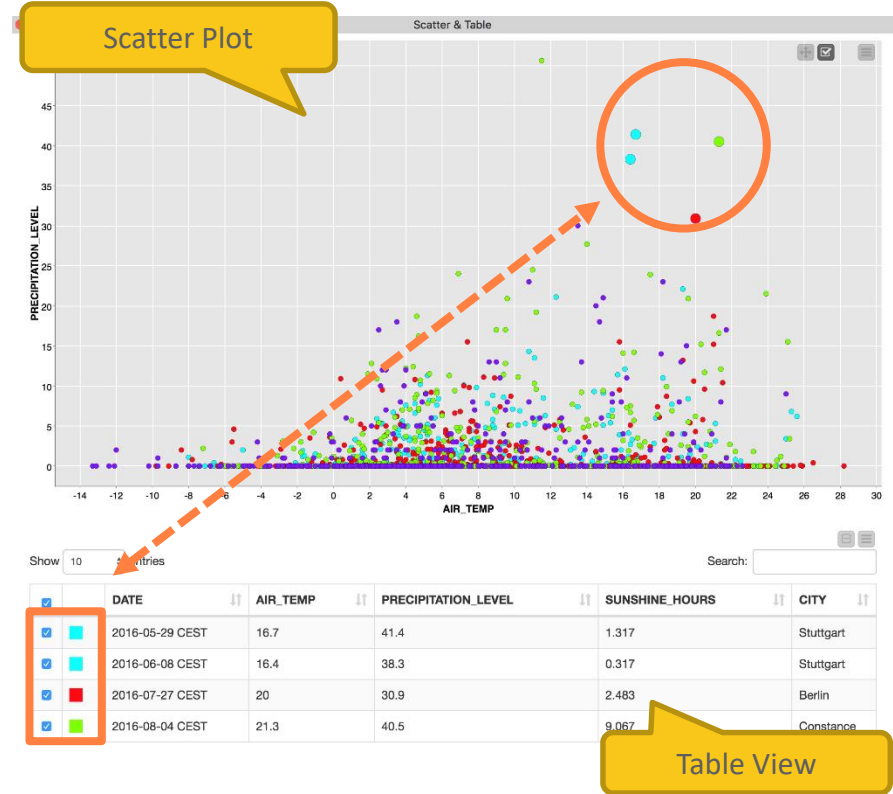
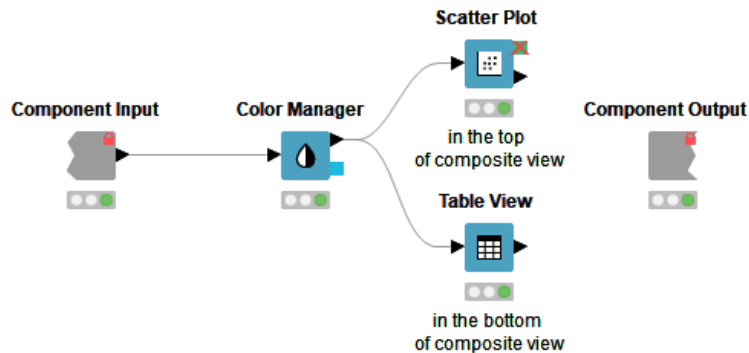
Interactivity allows you to select data points in views

- Selection is propagated to other views
- Highlight selected rows or filter them
- Click “Apply” to add column to data that indicates selection (true/false) for use in downstream nodes

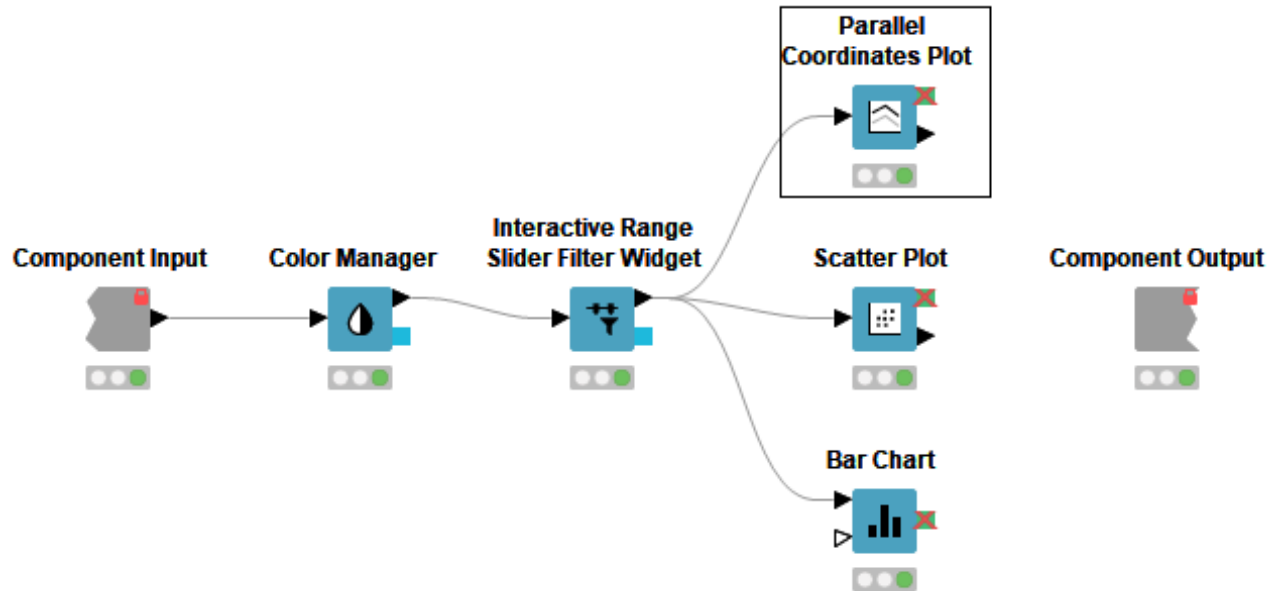


Components – Combined Views

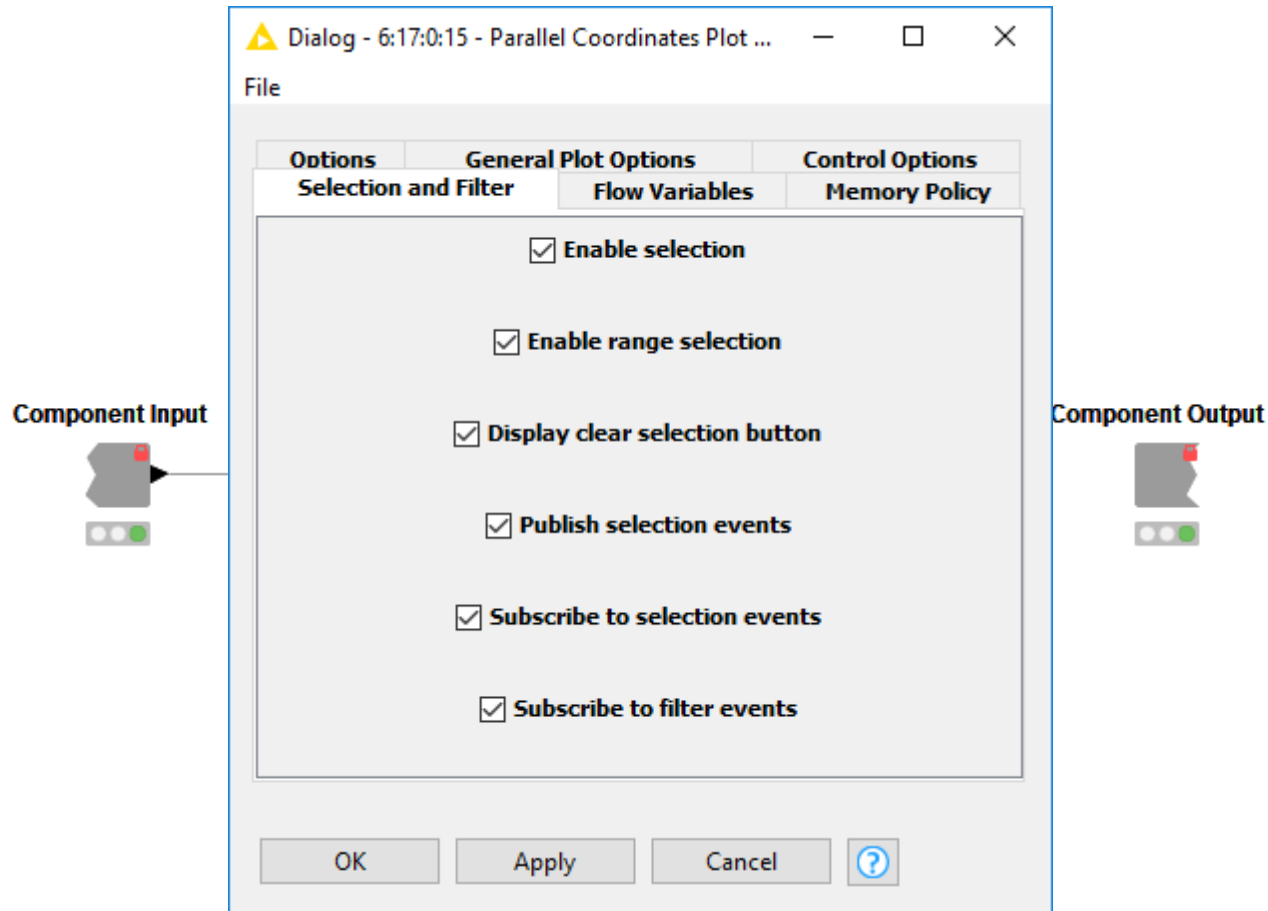
- Multiple JavaScript View nodes can be combined in Components
- Selections are transmitted to all other views
- Also for use on the KNIME WebPortal



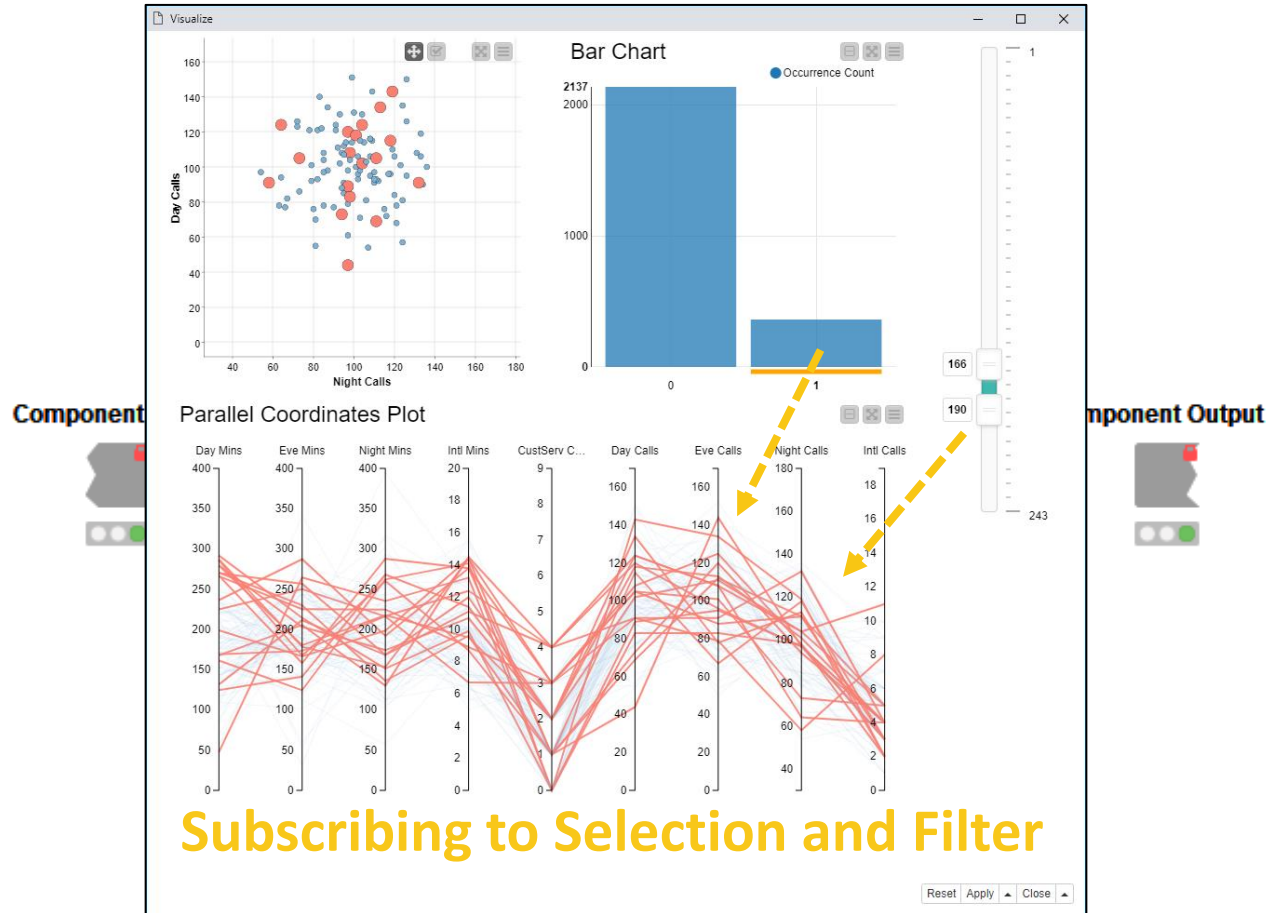
Interactivity across Charts: Selection and Filter Events



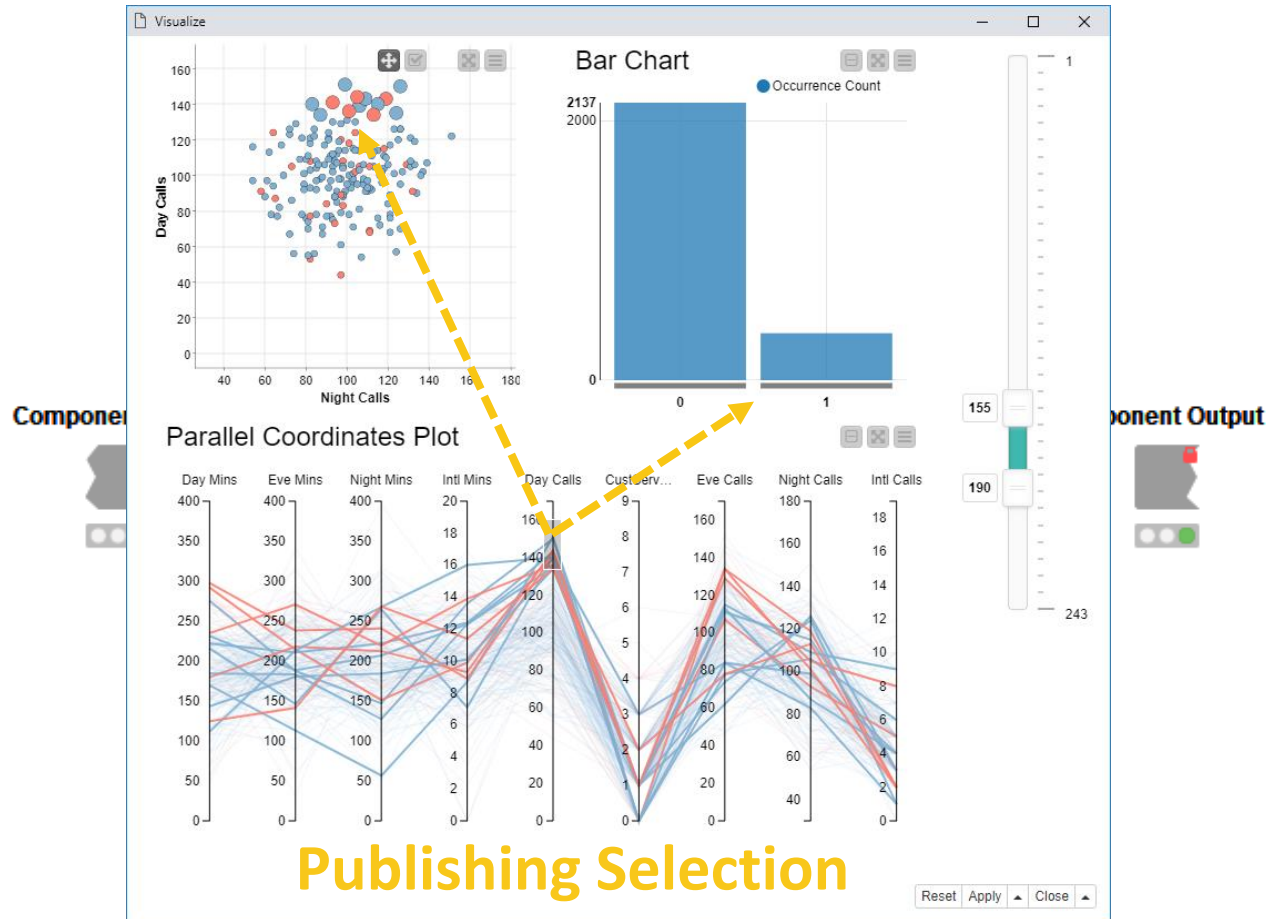
Interactivity across Charts: Selection and Filter Events



Interactivity across Charts: Selection and Filter Events

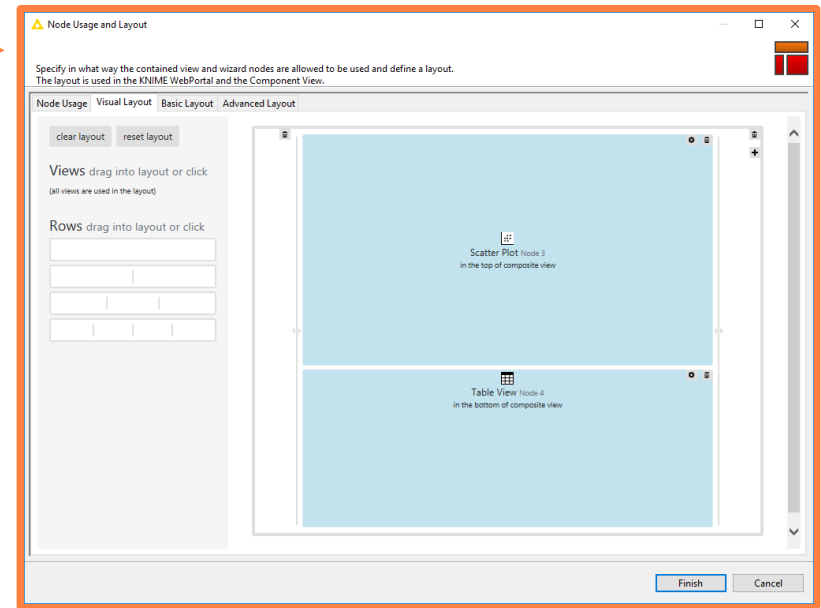
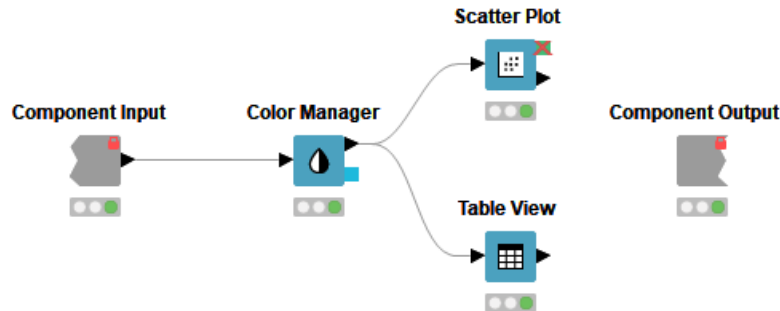
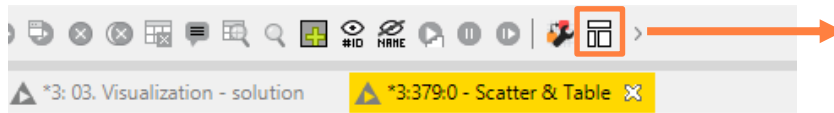


Interactivity across Charts: Selection and Filter Events



Configure Content and Views Layout

- Click layout button when inside Component to assign views to rows and columns
- Add views and rows via *drag&drop*
- Add columns using **+** buttons



Data Aggregation

Aggregation: Count

Product ID	Store	Category	# Ordered Items
P 1	Online	Clothing	2
P 2	Onsite	Home	3
P 3	Onsite	Clothing	1
P 4	Online	Clothing	5
P 5	Online	Electronics	7
P 6	Online	Electronics	5

Category	Online	Onsite
Clothing	2	1
Home	0	1
Electronics	2	0

Aggregation: Sum (# Ordered Items)

Category	Online	Onsite
Clothing	7	1
Home	0	3
Electronics	12	0

Solution: Pivoting Node

Data Aggregation

Product ID	Store	Category	# Ordered Items
P 1	Online	Clothing	2
P 2	Onsite	Home	3
P 3	Onsite	Clothing	1
P 4	Online	Clothing	5
P 5	Online	Electronics	7
P 6	Online	Electronics	5



Aggregation: Sum (# Ordered Items)

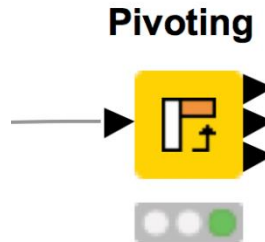
Category	Online	Onsite
Clothing	7	1
Home	0	3
Electronics	12	0

Pivoting Node: **Group** - **Pivot** - **Aggregate**

New Node: Pivoting

Performs pivoting on selected columns for grouping and pivoting

- Values of group columns become unique rows
- Values of the pivot columns become unique columns for each set of column combination together with each aggregation
- Many aggregation methods are provided (similar to GroupBy)



New Node: Pivoting

Groups ~ Rows

Pivots ~ Columns

Aggregation

Pivot table - 0:35 - Pivoting

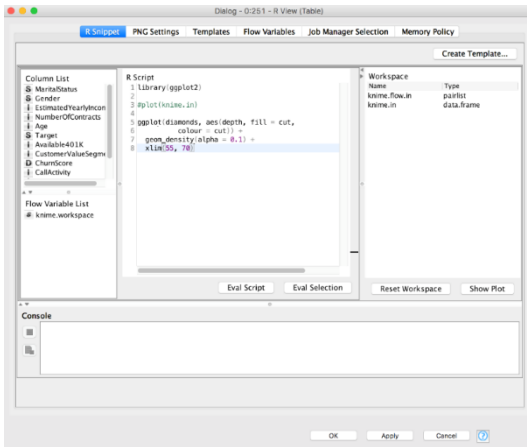
File Hilite Navigation View

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

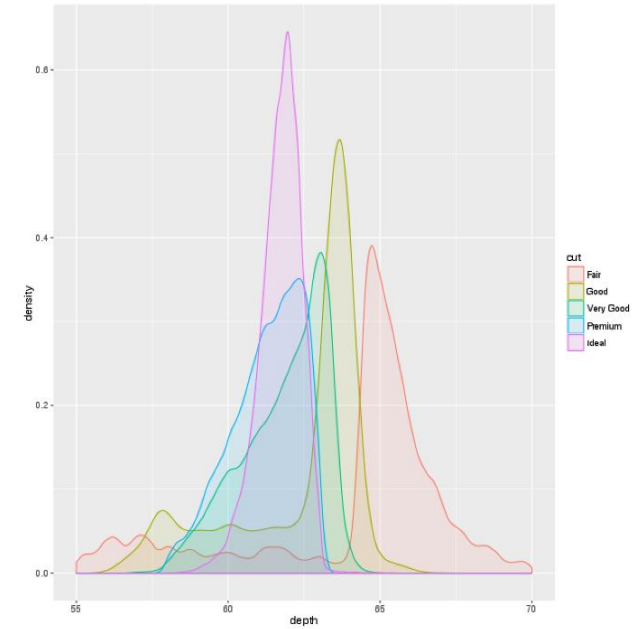
Row ID	Category	Online+Sum(OrderedItems)	Onsite+Sum(OrderedItems)
Row0	Clothing	11823	7604
Row1	Electronics	10754	6624
Row2	Home	7180	5109

Script-based View Nodes

- R View nodes for greater customizability
 - Use your favorite libraries, e.g. ggplot2
- If you prefer Python: Python View node
- For JS developers: Generic JavaScript View



R View (Table)



Visualization Exercise

Start with exercise: *Visualization*

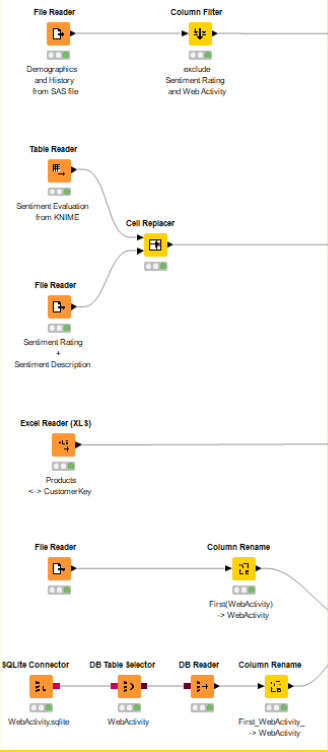
- Read *sales.csv* data
- Assign a different color to each product
- Plot BasketValue against BasketSize using the Scatter Plot node
- Show the total BasketValue by time and product in a Line Plot and a Stacked Area Chart (Use the Pivoting node to get the sum of sales by Quarter and Product!)
- Execute the *Fully Joined Data* metanode
- Show the number of customers in the different web activity categories in a Bar Chart
- Show the age distribution of the customers in a Histogram
- Create a composite view by combining the Bar Chart and Histogram
- Select one web activity class in the Bar Chart. Which age classes are represented in the selected web activity class?

Today's Example

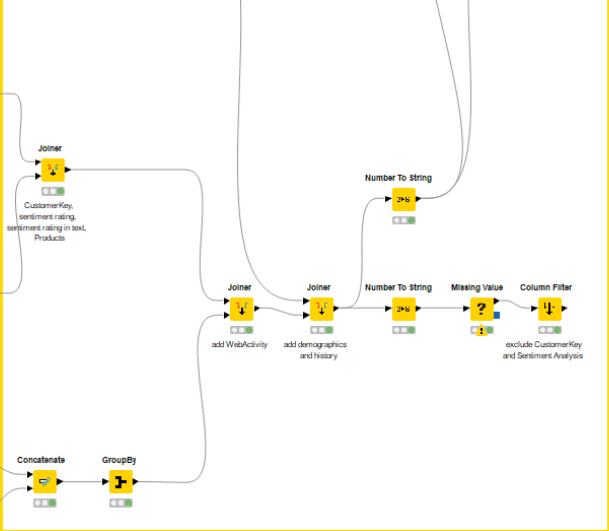
Final Workflow from the KNIME User Training

... and putting all those parts together, you get this final workflow.

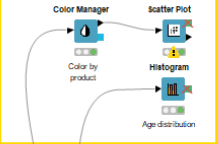
Data Reading



Data Manipulation and Aggregation



Visualization



Training Predictive Models

Data Export and Reporting

Data Mining

Partition, Learn, Predict, Score

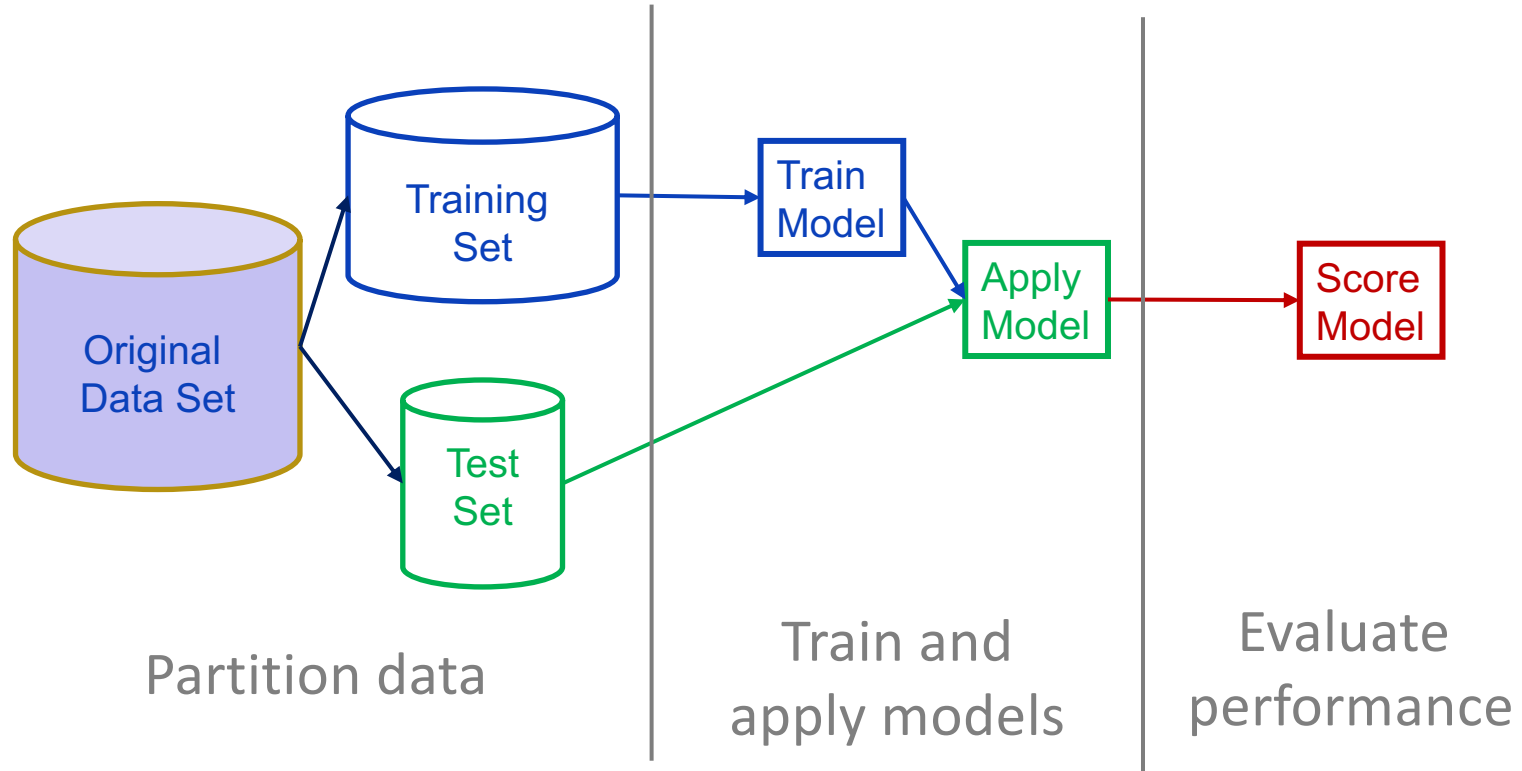


Data Mining Strategies

Example Applications:

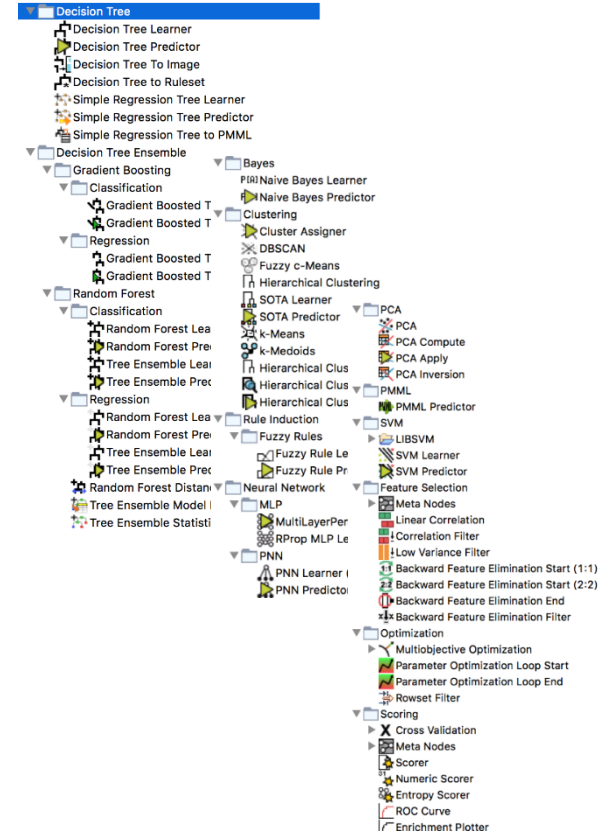
- Anomaly Detection (fraud, predictive maintenance)
- Association Rule Learning (market basket analysis)
- Clustering (market segmentation)
- Classification (next best offer, churn preventions)
- Regression (trend estimation)

Data Mining: Process Overview



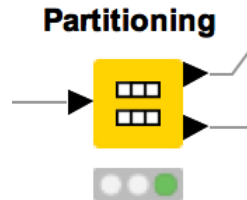
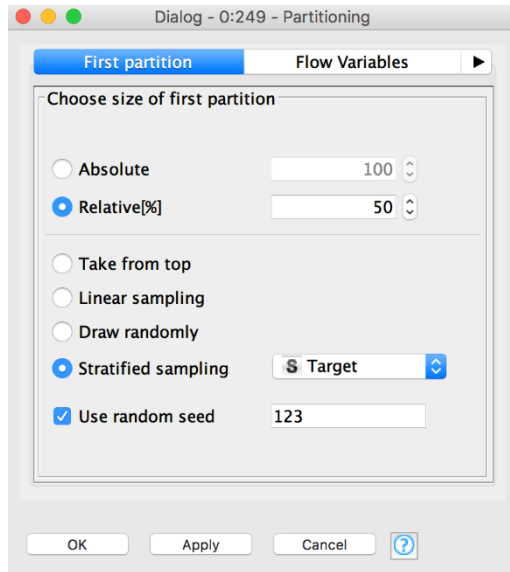
Data Mining in KNIME

- KNIME has many modeling tools!
 - Decision tree, random forest, SVM, regression, neural networks, clustering, ...
 - and integrations with other libraries: R, Python, H2O, WEKA, libSVM, etc.
- And many model evaluation nodes
 - ROC, standard, numeric and entropy scorers
 - Feature elimination
 - Cross validation



New Node: Partitioning

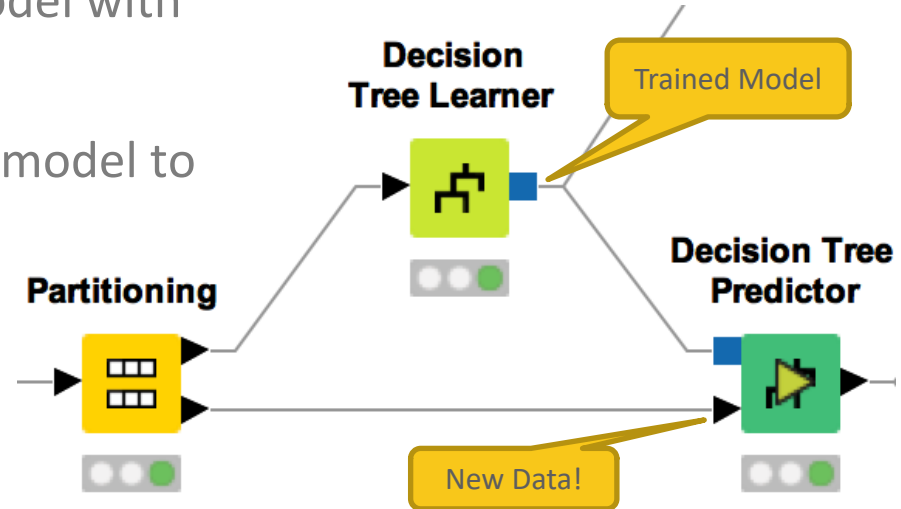
- Use to split data into training and evaluation sets
 - Partition by count (e.g. 10 rows) or fraction (e.g. 10%)
 - Sample by a variety of methods; random, linear, stratified



The first screenshot shows the 'First partition (as defined in dialog) - 0:249 - Partitioning' window. It displays a table with 13 columns: Row ID, \$ Marita..., \$ Gender, \$ Estim..., \$ Numb..., and Age. The table contains 5775 rows, with the first 50 rows (Row0 to Row49) highlighted in black. The second screenshot shows the 'Second partition (remaining rows) - 0:249 - Partitioning' window, displaying the same table structure but with the remaining rows (Row50 to Row5774) highlighted in black.

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.

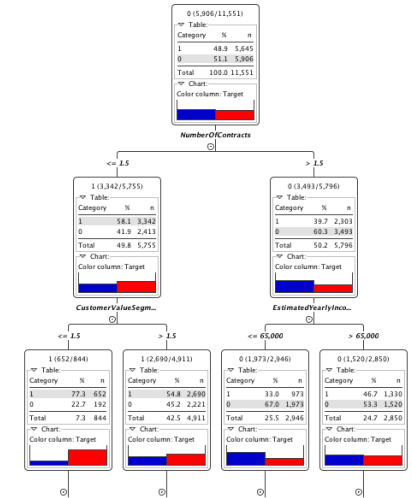
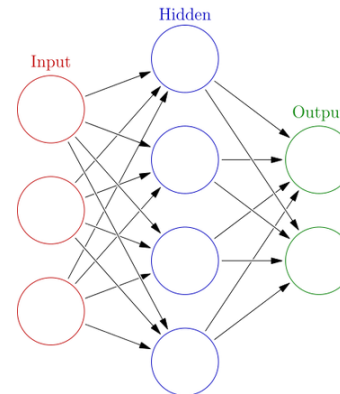


Classification

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%

Target Column

- Target column contains values that are predicted by the classification model
- Binomial target values are often encoded to 1 and 0

Application	Target Column	Target Values
Churn analysis	Churn	Yes/No or 1/0
Chemical activity	Active	Yes/No or 1/0
Spam Detection	Spam	Yes/No or 1/0
Optical Character Recognition	Character	A-Z

Output data - 0:311 - Column Resorter

File Hilite Navigation View

Table "default" - Rows: 5776 Spec - Columns: 17 Properties Flow Variables

R...	I CustomerKey	S Marital...	S Gender	S Target	S Prediction (Target)
...	11001	S	M	1	0
...	11002	M	M	1	0
...	11003	S	F	1	1
...	11004	S	F	1	0
...	11005	S	M	1	1
...	11006	S	F	1	1
...	11008	S	F	1	0
...	11011	M	M	1	0
...	11012	M	F	0	1
...	11016	M	M	1	1

KNIME's Decision Tree

J.R. Quinlan, "C4.5 Programs for machine learning"

J. Shafer, R. Agrawal, M. Mehta, "SPRINT: A Scalable Parallel Classifier for Data Mining"

- C4.5 builds a tree from a set of training data using the concept of information entropy.
- At each node of the tree, the attribute of the data with the highest **normalized information gain** (difference in entropy) is chosen to split the data.
- The C4.5 algorithm then recurses on the smaller sub lists.

New Node: 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

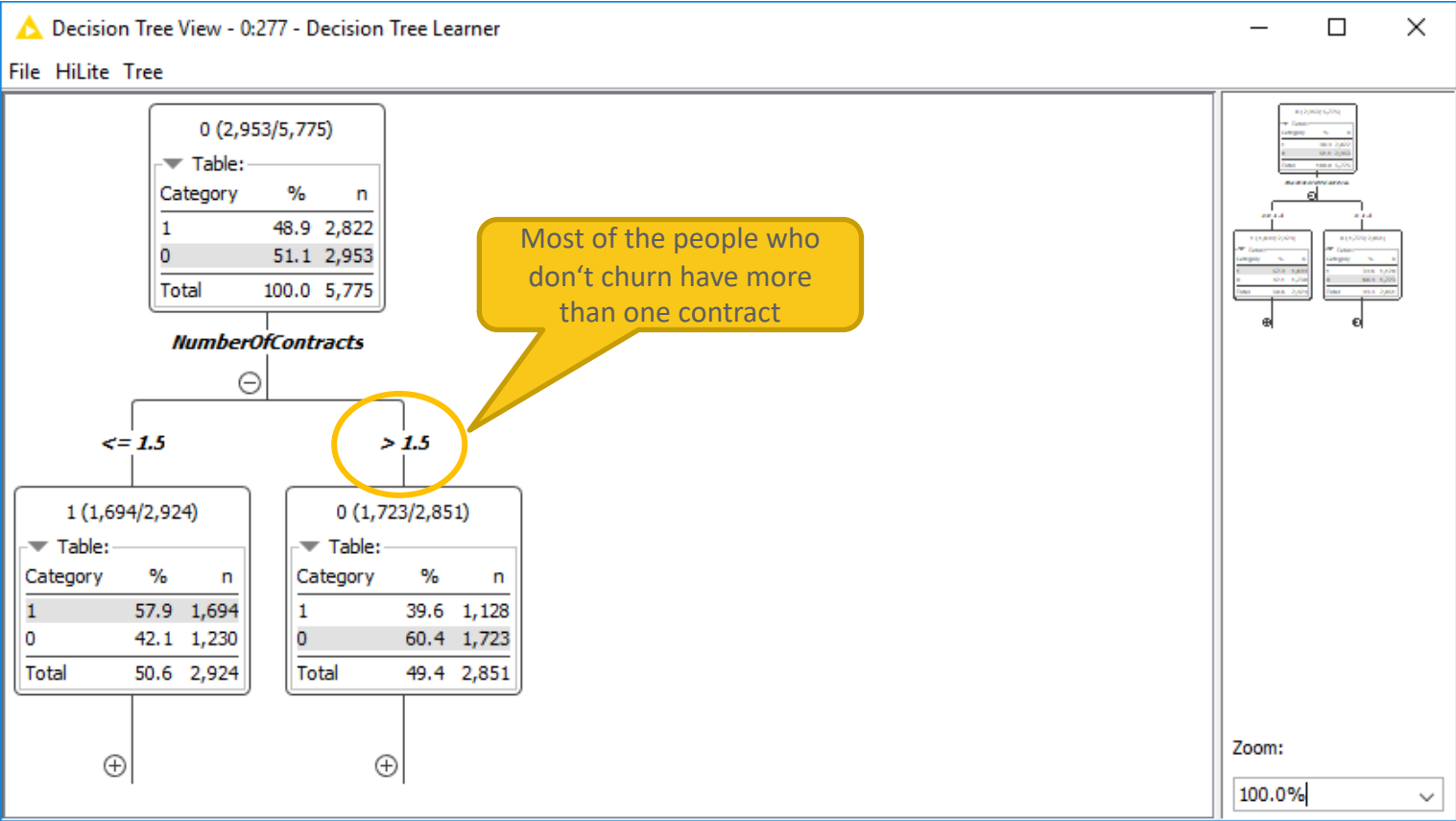
☐ Binary nominal splits

Max #nominal

☐ Filter invalid attribute values in child nodes

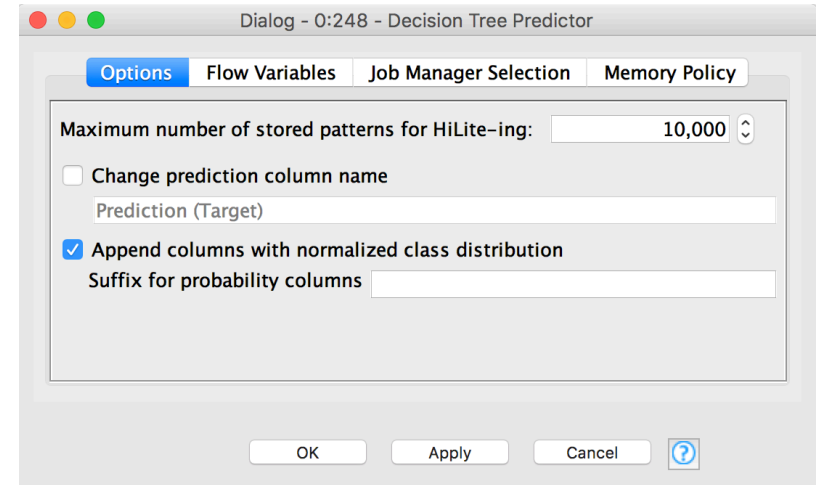
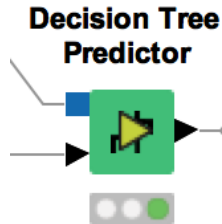
OK Apply Cancel ?

Decision Tree View



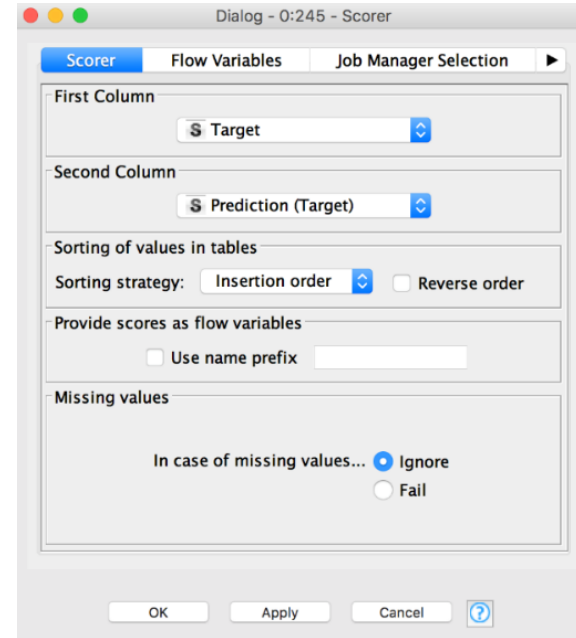
New Node: Decision Tree Predictor

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



New Node: Scorer

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



New Node: 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 (κ) 0.477

An accuracy statistics table provides a detailed analysis of model quality

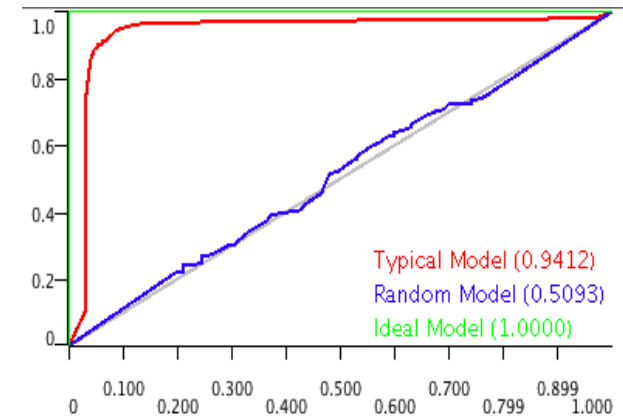
Accuracy statistics - 0:297 - Scorer											
File Hilite Navigation View											
Table "default" - Rows: 3 Spec - Columns: 11 Properties Flow Variables											
Row ID	TruePositives	FalsePositives	TrueNegatives	FalseNegatives	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy	Cohen's 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

Confusion Matrix

	Predicted class POSITIVE (churn)	Predicted class NEGATIVE (no churn)
Actual class POSITIVE (churn)	TRUE POSITIVE (TP) 2073	FALSE NEGATIVE (FN) 750
Actual class NEGATIVE (no churn)	FALSE POSITIVE (FP) 759	TRUE NEGATIVE (TN) 2193

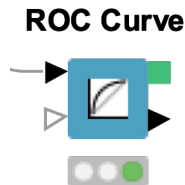
Receiver Operating Characteristics

- Sort by confidence in target class
- Plot true positive rate vs false positive rate
- Ideal models achieve 100% TPR with 0% FPR
- Area under the curve indicates model quality
 - (1=ideal model, 0.5 = random outcome)



New Node: ROC Curve

- Requires individual class probabilities from a preceding predictor
- User must define:
 1. Original class column
 2. Positive class value
 3. Probability for the selected positive class value for one or multiple models



Dialog - 0:278 - ROC Curve (JavaScript)

ROC Curve Settings General Plot Options Axis Configuration View Controls Flow Variables

Class column: Target

Positive class value: 1

Limit data points for each curve to: 12,000

Columns containing the positive class probabilities

☒ Manual Selection ☐ Wildcard/Regex Selection

Exclude

Filter

- EstimatedYearlyIncome
- NumberOfContracts
- Age
- Available401K
- CustomerValueSegment
- ChurnScore
- CallActivity

☐ Enforce exclusion

Include

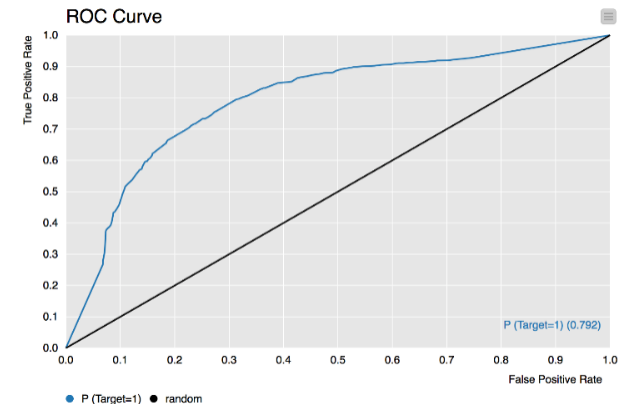
Filter

- P (Target=1)

☒ Enforce inclusion

☐ Ignore missing values

OK Apply Cancel ?



Data Mining Exercise, Activity I

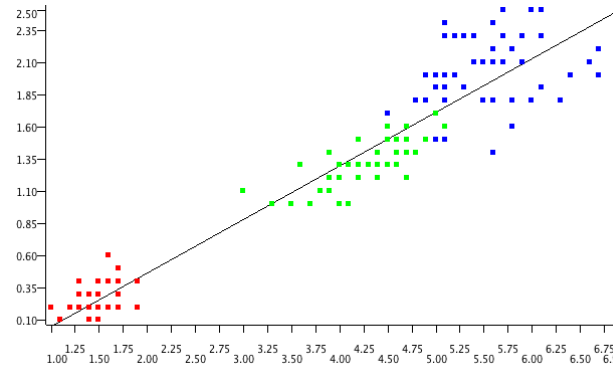
Start with exercise: *Data Mining, Activity I*:

- Partition the fully joined data into a training and test set (50%, Stratified Sampling on Target)
- Train a decision tree on the training set to predict Target
- Use the trained model to predict Target in the test set
- What is the overall accuracy of your model?
- Optional: Evaluate the accuracy and robustness of the model with the ROC Curve node

Regression

Predict *numeric* outcomes on existing data (supervised)

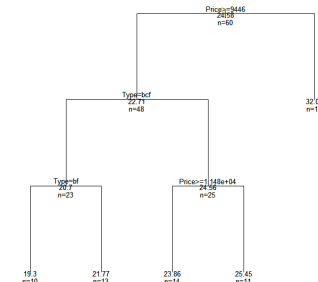
- Applications
 - Forecasting
 - Quantitative Analysis
- Methods
 - Linear
 - Polynomial
 - Regression Trees
 - Partial Least Squares



Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P> t
Petal.Length	0.4158	0.0096	43.3872	0.0
Intercept	-0.3631	0.0398	-9.1312	4.44E-16

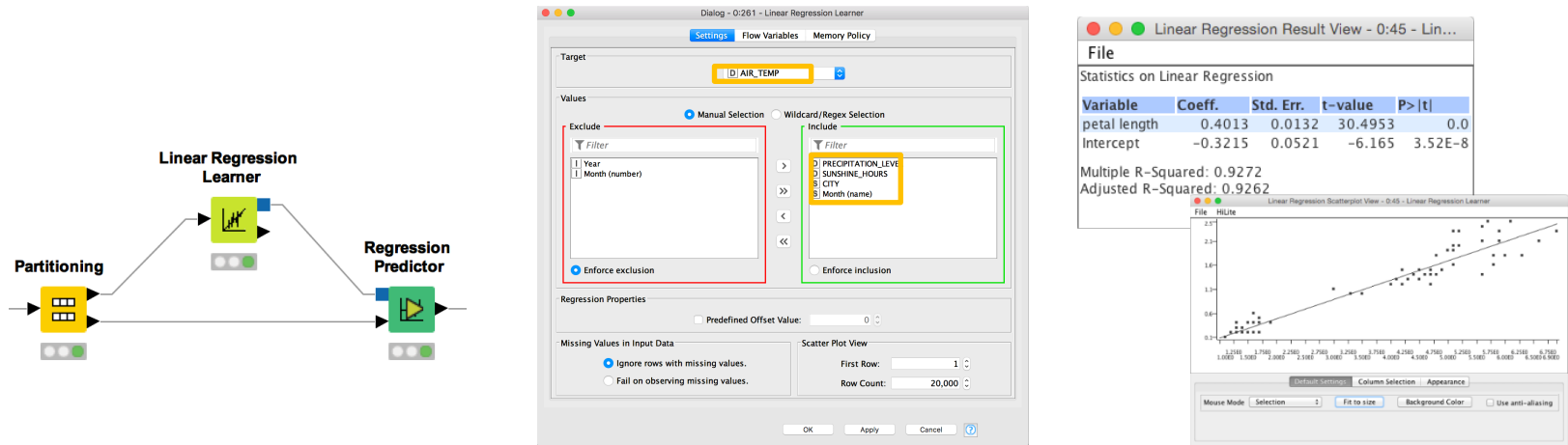
Multiple R-Squared: 0.9271
Adjusted R-Squared: 0.9266



New Nodes: Linear Regression Learner & Regression Predictor

A linear model relating a dependent variable to 1 or more independent variables

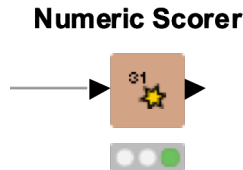
- Model coefficients provided in 2nd output port
- Also available: Polynomial and Tree Ensemble Regression nodes



New Node: Numeric Scorer

Similar to scorer node, but for nodes with *numeric* predictions (e.g. linear/polynomial regression)

- Compare dependent variable values to predicted values to evaluate goodness of fit.
- Report R^2 , MAE, MSE, RMSE etc.



Dialog - 0:298 - Numeric Scorer

File

Options Flow Variables Job Manager Selection Memory Policy

Reference column [D] AIR_TEMP

Predicted column [D] Prediction (AIR_TEMP)

Output column

☐ Change column name

Output column name Prediction (AIR_TEMP)

Provide scores as flow variables

Prefix of flow variables

☐ Output scores as flow variables

OK Apply Cancel ?

Statistics - 0:298 - Numeric Scorer

File Hilite Navigation View

Table "Scores" - Rows: 6 Spec - Column: 1 Properties Flow Variables

Row ID	[D] Prediction (AIR_TEMP)
R^2	0.333
mean absolute error	3.574
mean squared error	21.329
root mean squared error	4.618
mean signed difference	1.048
mean absolute percentage error	NaN

Data Mining Exercise, Activity II

Start with exercise: *Data Mining, Activity II*:

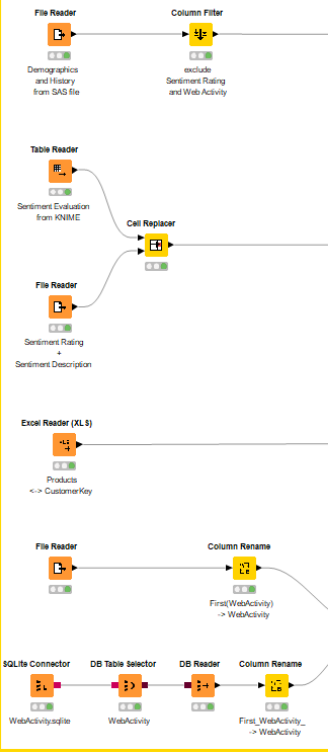
- Read *weather.table* data
- Split the data into rows up to 2016 (training set) and rows from 2017 on (test set)
- Train a linear regression model that predicts the AIR_TEMP as a function of all other features in the dataset
- Use the model to predict the temperature in 2017 and evaluate the model with the Numeric Scorer node
- Optional:
 - Calculate the mean temperature per month in the training data
 - Join the mean temperature per month to the test set
 - Use the Numeric Scorer to see if the average monthly temperature provides a better prediction than the Linear Regression model

Today's Example

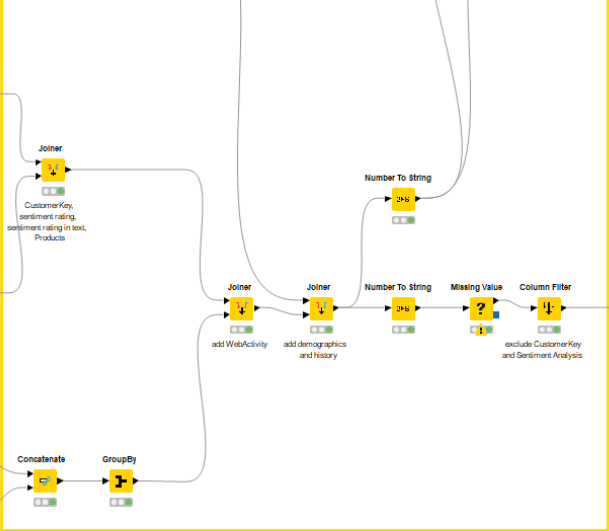
Final Workflow from the KNIME User Training

... and putting all those parts together, you get this final workflow.

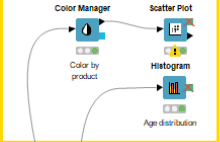
Data Reading



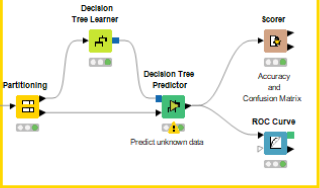
Data Manipulation and Aggregation



Visualization



Training Predictive Models

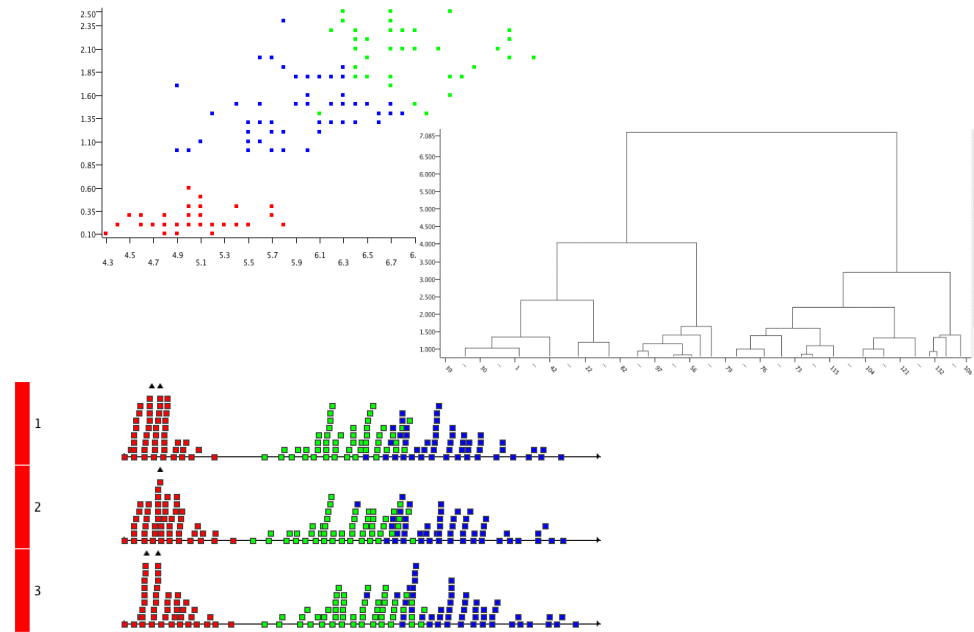


Data Export and Reporting

Clustering

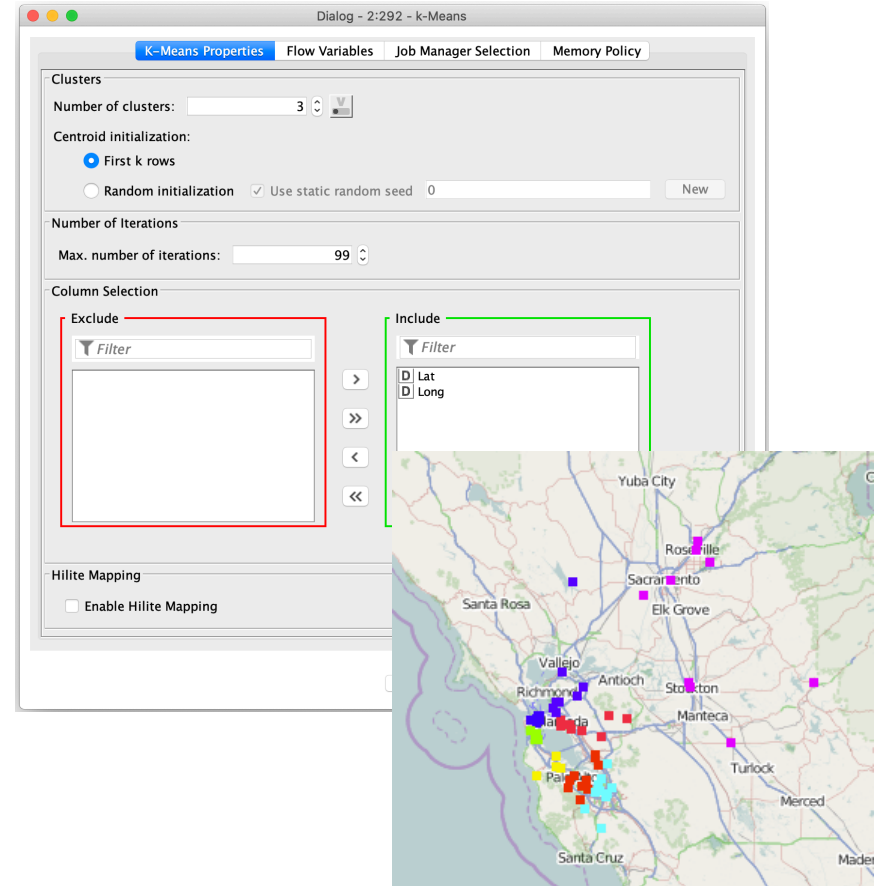
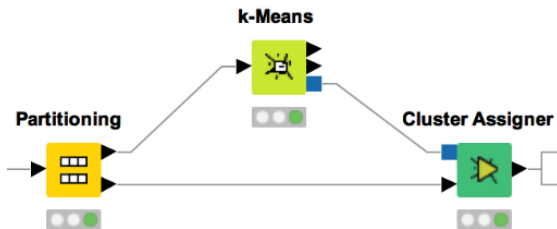
Discover hidden structure in **unlabeled** data (unsupervised)

- Applications
 - Market Segmentation
 - Diversity picking
- Methods
 - K-means/medoids
 - Hierarchical
 - DBScan
 - OPTICS
 - Neighbourgrams



New Nodes: k-Means Clustering

- Looks at n observations to define the means for k clusters.
- Each observation is then assigned to its closest cluster center.
- You must provide k .



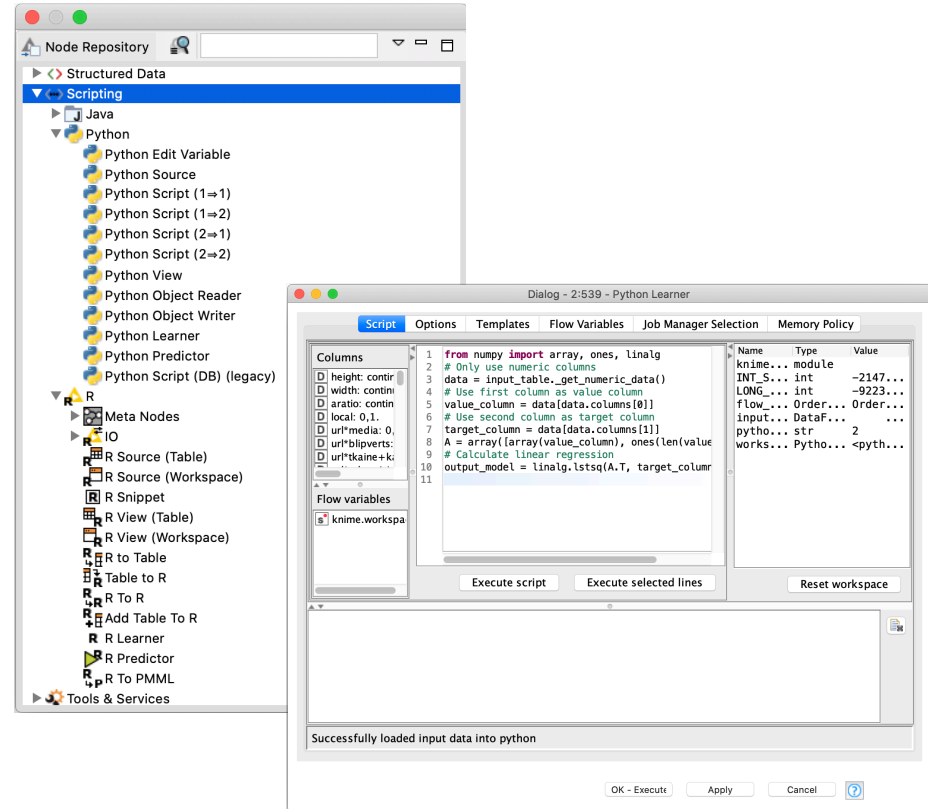
Data Mining Exercise, Activity III

Start with exercise: *Data Mining, Activity III*

- Read *location_data.table* data
- Filter the data to entries from California (region_code = CA)
- Perform k-means clustering with k=3. Use only latitude and longitude for clustering.
- Optional: plot latitude and longitude in a view (OSM Map or Scatter Plot) and use the view to visually optimize k

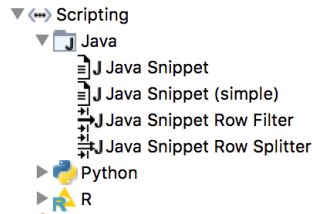
Scripting Integrations: R and Python

- Run R or Python code in KNIME Analytics Platform
- Works with existing Python and R installations
- Syntax highlighting support
- Different nodes for many tasks, e.g training a model available in Python

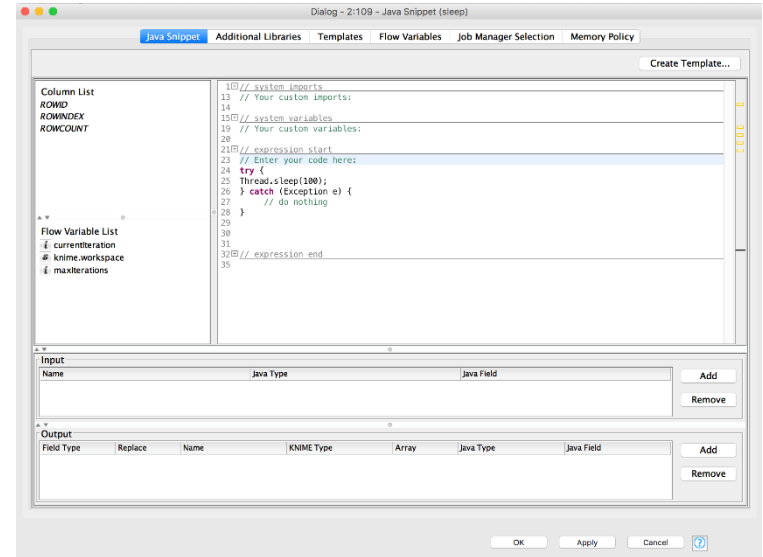
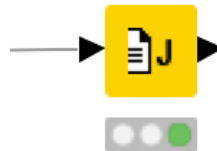


Java Snippet

- Fastest running scripting node in KNIME
- Syntax highlighting, auto completion, error checking
- Templates allow you to save scripts for later re-use
- Import custom libraries



Java Snippet



Exporting Data & Deployment

Exporting Data

After an analysis is completed, what next?

- Write results to a file
- Create/update a database
- Save the model for use elsewhere
- Generate a rich report
- Deploy via KNIME WebPortal
- Deploy via workflow as RESTful web service

Input/Output in Deployment

Input

- File (CSV, Table, XLS, ...)
- Database
- JSON for REST API

Output

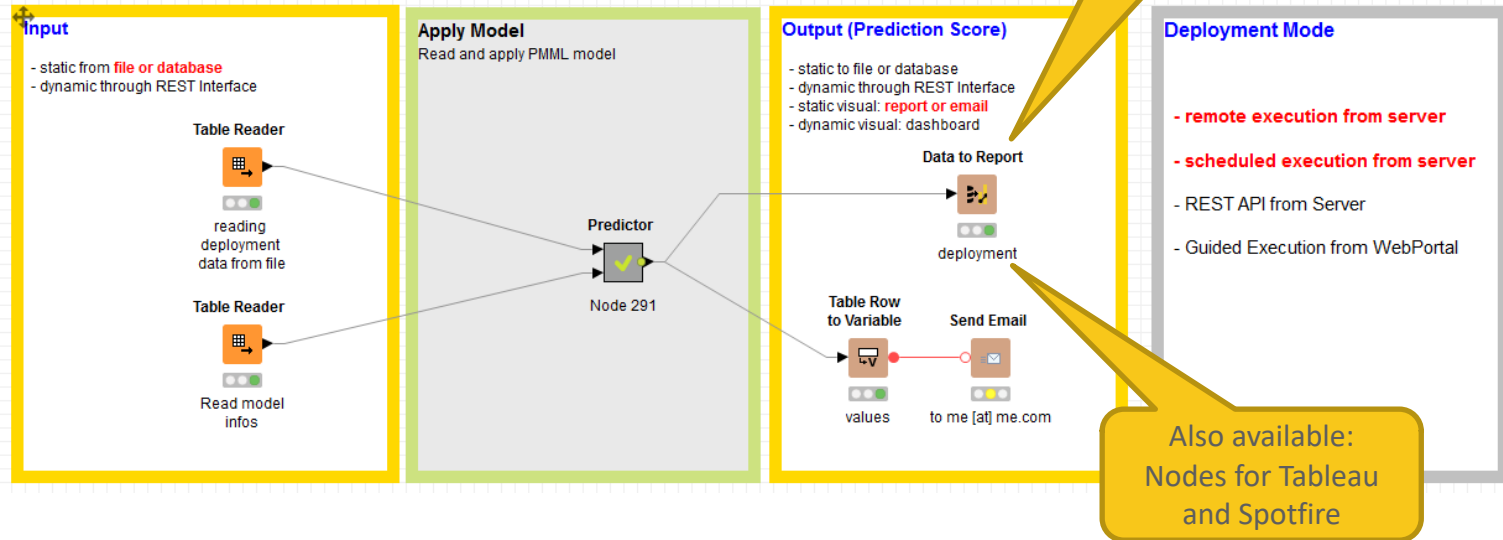
- Report (BIRT, Tableau, Spotfire, PowerBI)
- Email
- File (CSV, Table, XLS, ...)
- WebPortal

To Report / Email

Model Deployment with final report (Scheduling)

This workflow:

- reads new unseen data from file (.table format),
- scores the data with the available current model,
- appends model prediction and probabilities to original data
- produces a report (BIRT here) with table, bar chart, title, etc ... Report can be exported as .docx, html, pptx, .ps, .pdf, etc ...



To File / Database

Model Deployment File to Database (Scheduling)

This workflow:

- reads new unseen data from file (.table format),
- scores the data with the available current model,
- appends model prediction and probabilities to original data
- writes results to database

Input

- static from **file or database**
- dynamic through REST Interface

Table Reader



reading
deployment
data from file

Table Reader



Read model
infos

Apply Model

Read and apply PMML model

Predictor



Node 291

Output (Prediction Score)

- static to **file or database**
- dynamic through REST Interface
- static visual : report or email
- dynamic visual: dashboard

Database Writer



table
DeploymentData
WithPredictions



connect to DB

Deployment Mode

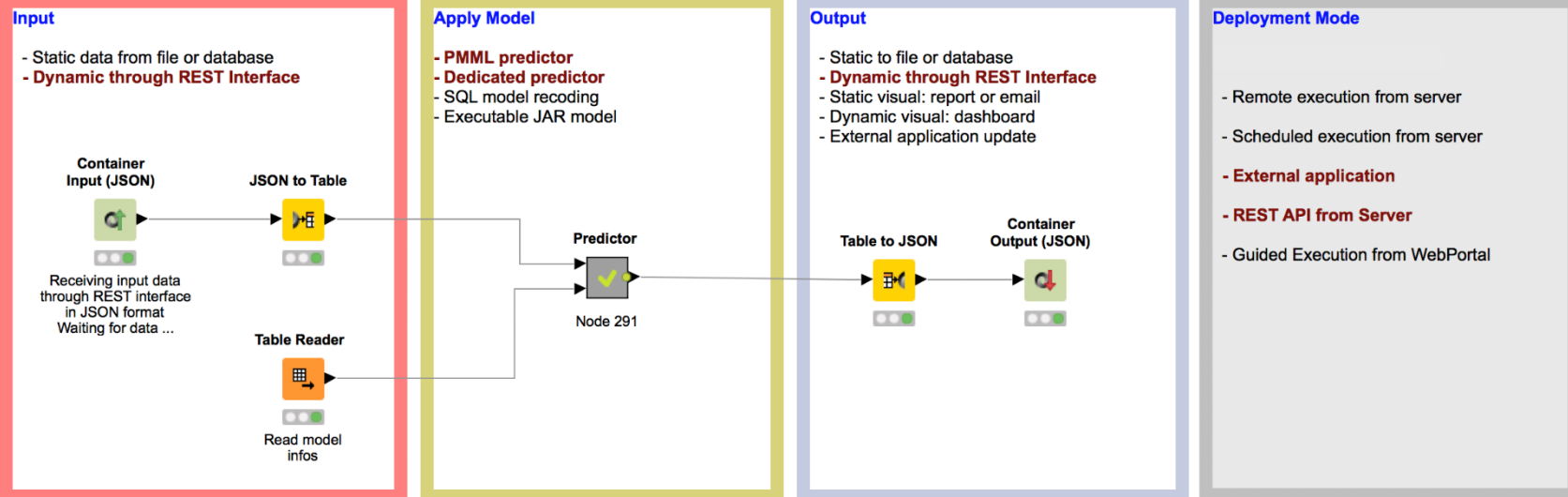
- **remote execution from server**
- **scheduled execution from server**
- REST API from Server
- Guided Execution from WebPortal

REST API (Available on KNIME Server)

Model Deployment as REST API

This workflow:

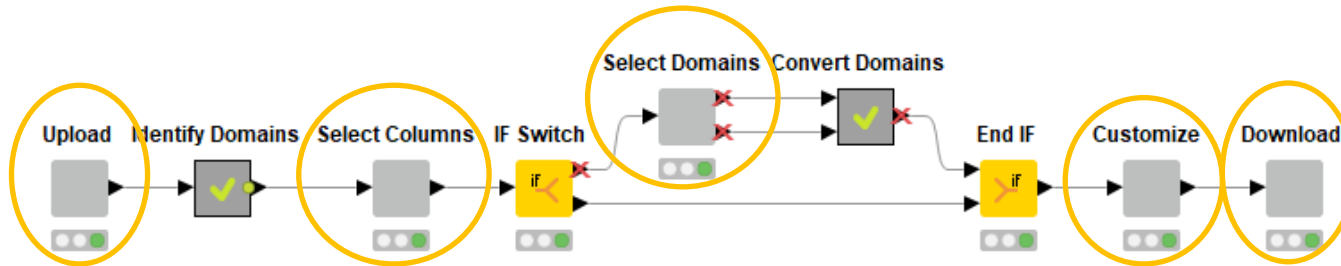
- receives new unseen data via REST interface (JSON format),
- scores the data with the available current model,
- appends model prediction and probabilities to original data,
- makes results available at the output REST interface.



To Dashboard on WebPortal

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

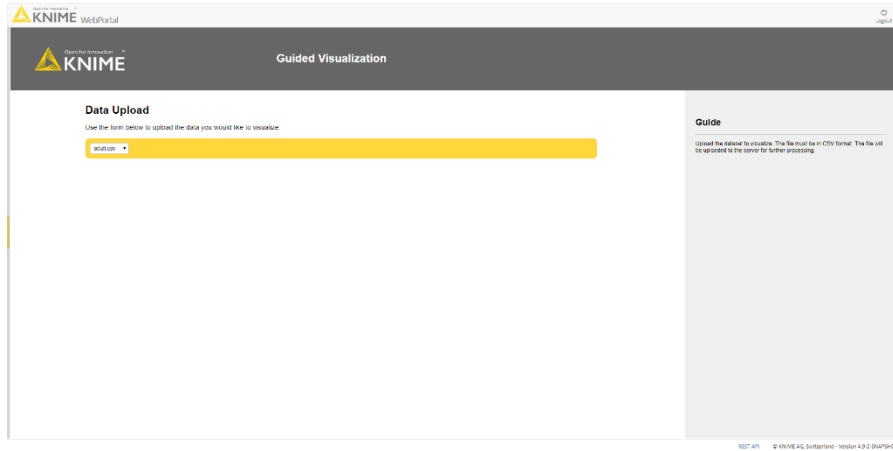
Step 2
Select Columns

Step 3
Customize
Column Domains

Step 4
Interactive View

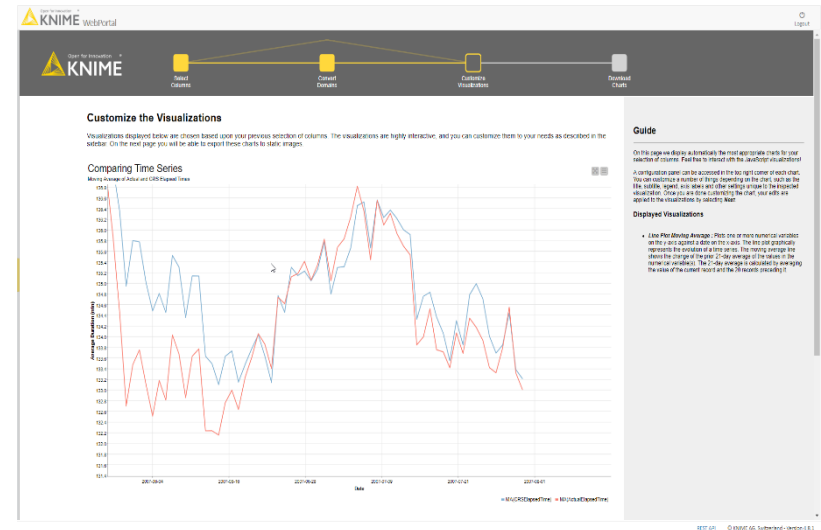
Step 5
Download Image

Workflow on KNIME WebPortal



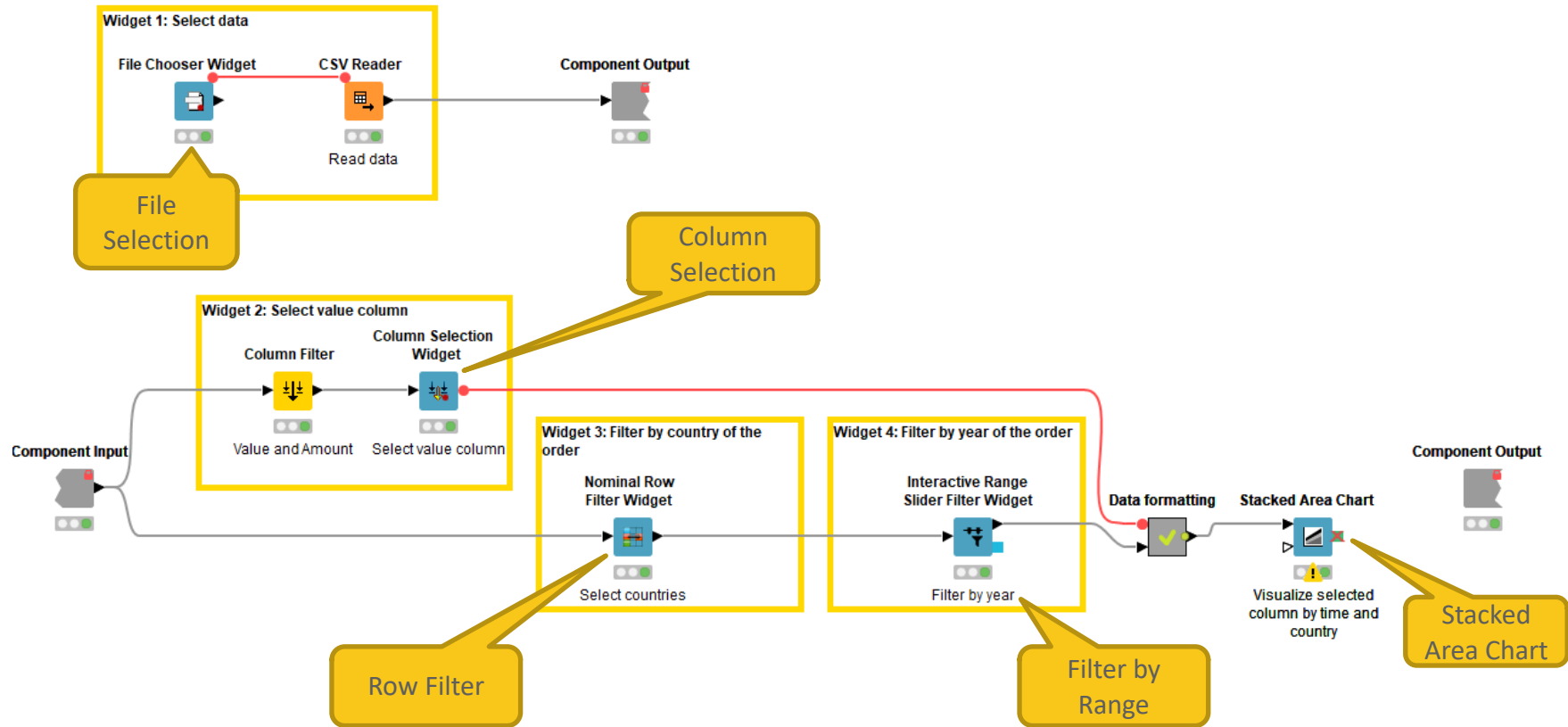
**WebPortal Page
(Step 1)
Upload File**

**Available in
KNIME Server**



**WebPortal Page
(Step 4)
Interactive View**

Components to Produce Dashboard on Web Page

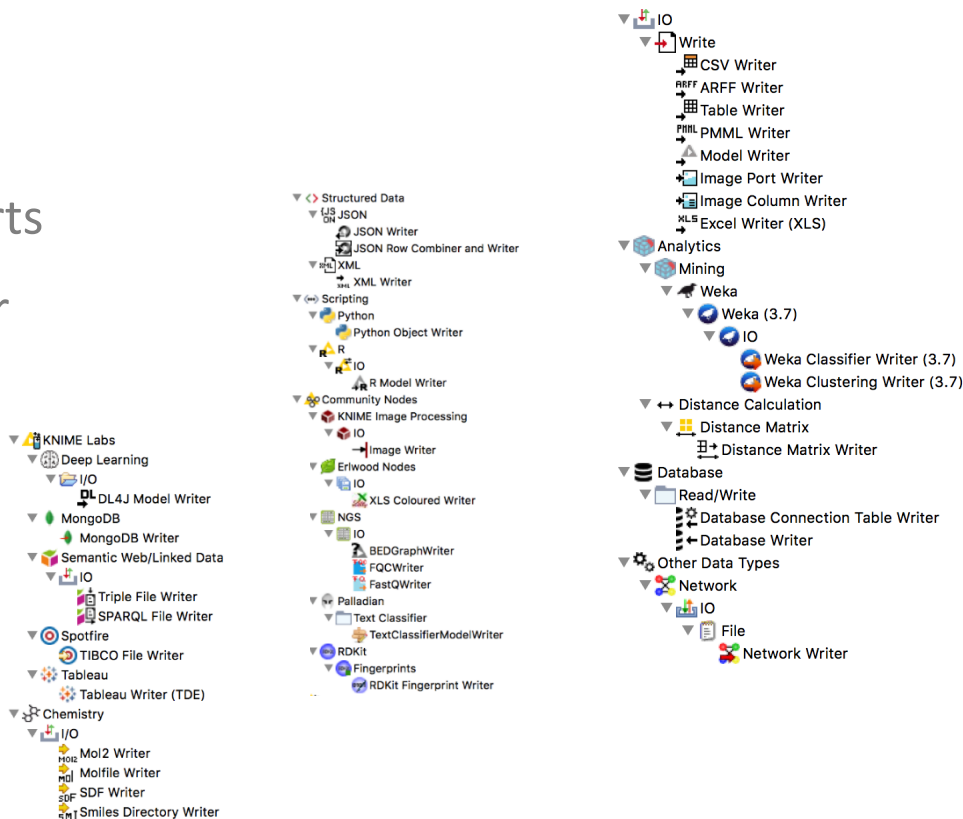
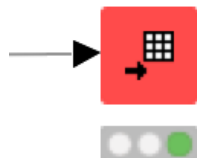


Data Export Nodes

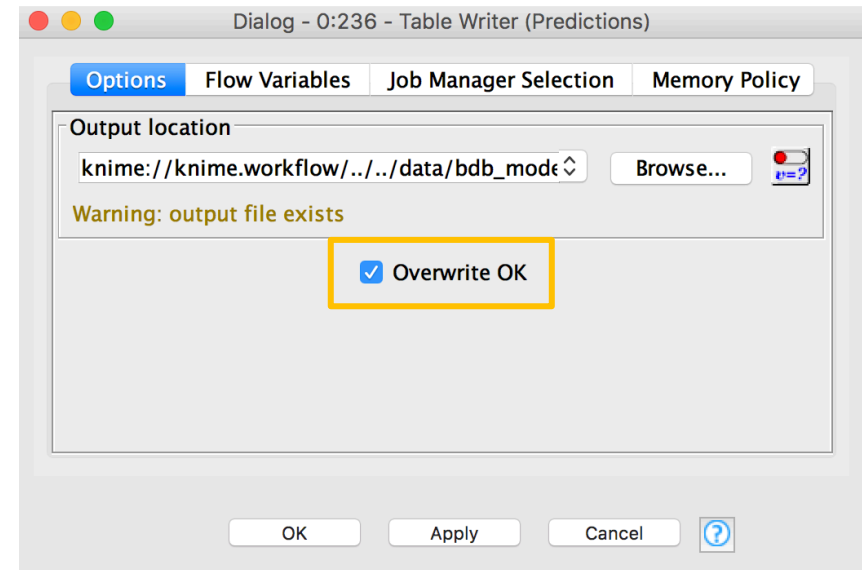
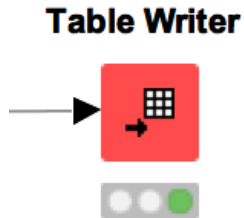
Typically characterized by:

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

Table Writer

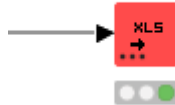


New Node: Table Writer



New Node: XLS Writer

Excel Writer (XLS)



Dialog - 3:260 - Excel Writer (XLS)

File

Settings | Flow Variables | Job Manager Selection | Memory Policy

Output location:

Write to: Relative to knime.workflow

File: ../data/example.xlsx Browse...

Scanning...

☐ Overwrite existing file

☐ Open file after execution

Sheet name

Name of the sheet: default

Add names and IDs

☐ add column headers

☐ add row ids

Missing value pattern

☐ For missing values write:

Layout

☐ Autosize columns

☒ Portrait ☐ Landscape US Letter 8 1/2 x 11 in

☒ Manual Selection ☐ Wildcard/Regex Selection ☐ Type Selection

Exclude

Filter

No columns in this list

☒ Enforce exclusion

Include

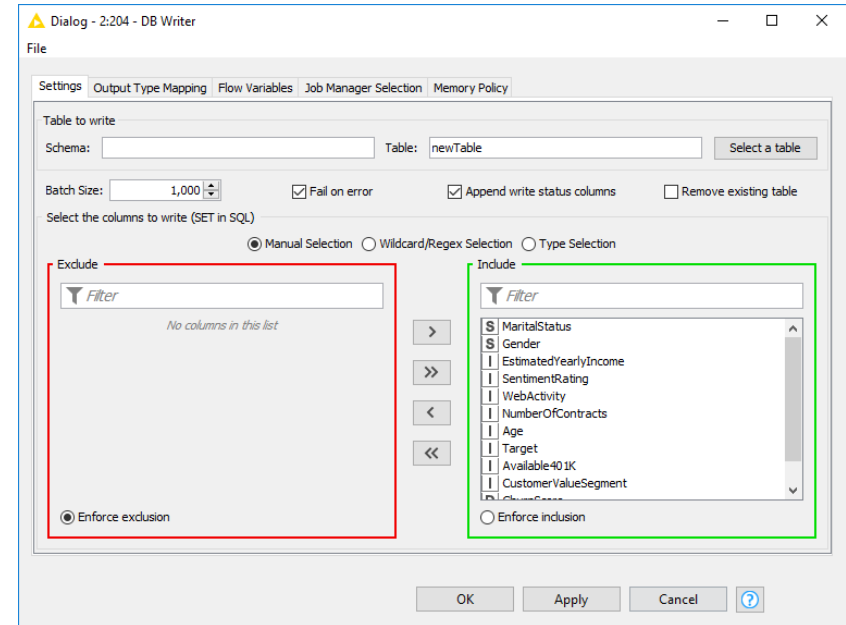
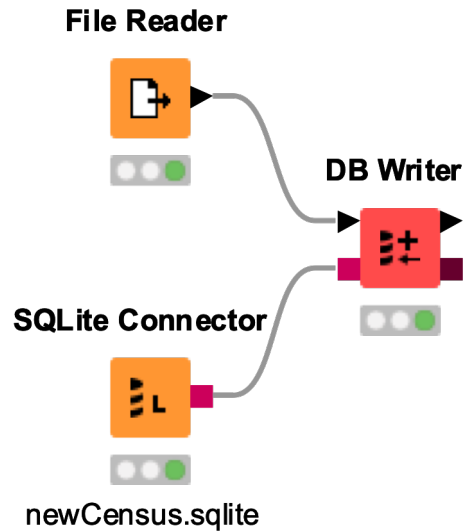
Filter

- ☒ CustomerKey
- ☒ MaritalStatus
- ☒ Gender
- ☒ EstimatedYearlyIncome
- ☒ NumberOfContracts
- ☒ Age
- ☒ Target
- ☒ Available401K
- ☒ CustomerValueSegment

☐ Enforce inclusion

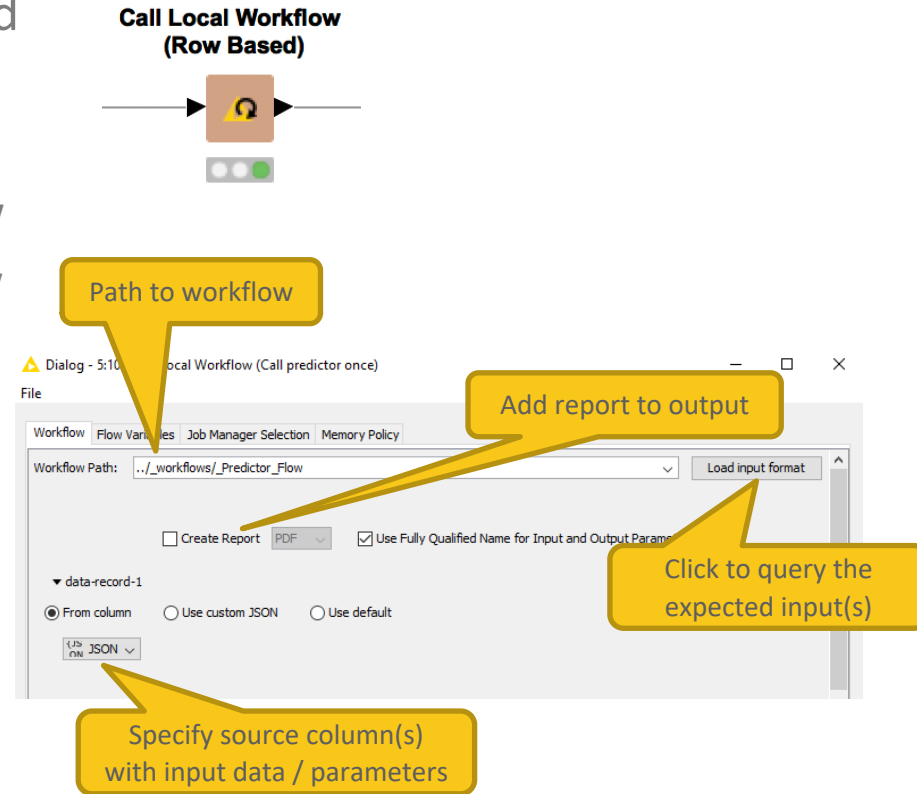
OK Apply Cancel ?

New Node: Database Writer



Automation: Call Local Workflow

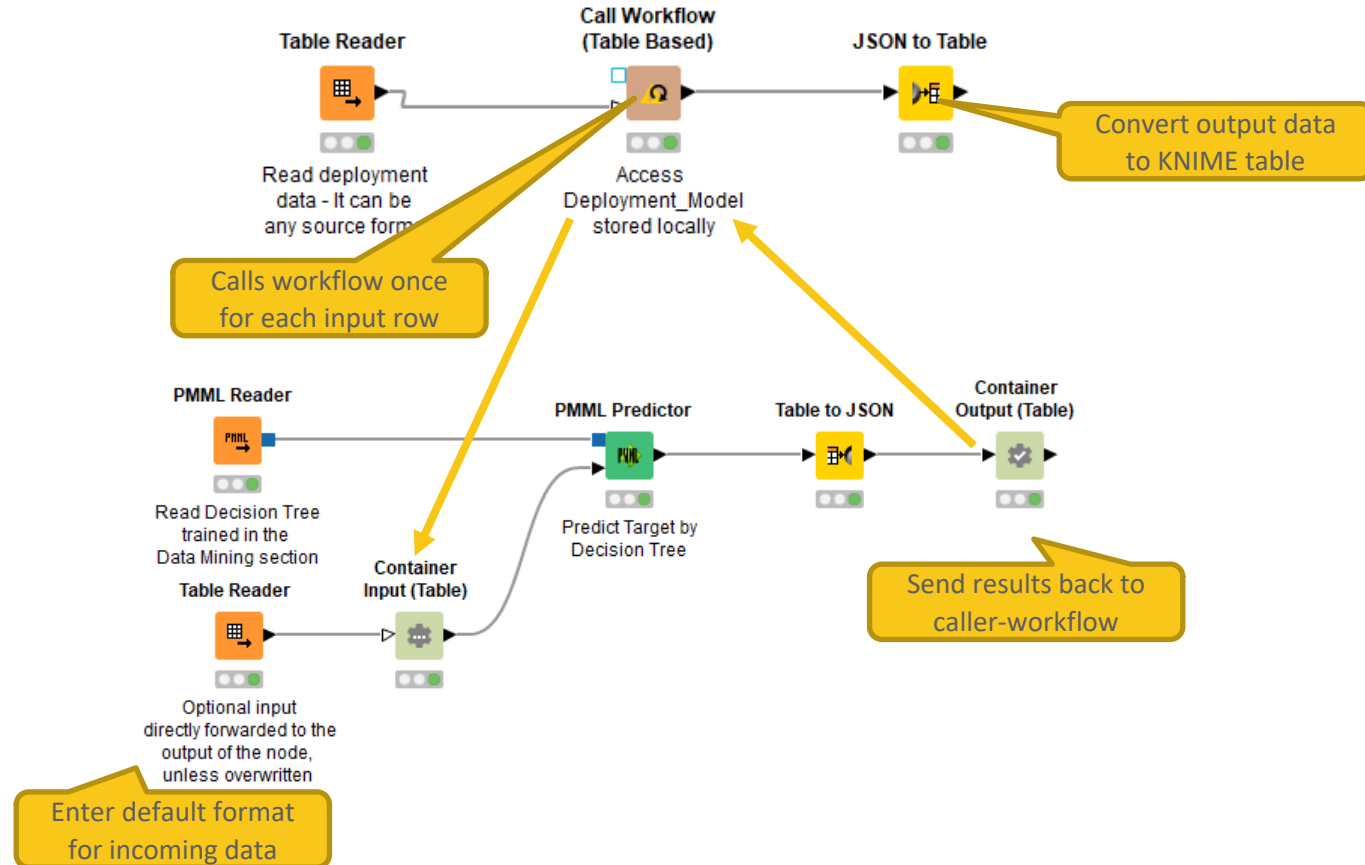
- Use Call Local Workflow node to send data and parameters to other workflows and trigger execution
 - Send results back to caller-workflow
 - Include report from called workflow
- Create modular workflows
 - E.g. separate workflows for ETL and prediction
- Alternative: Call Remote Workflow
 - Trigger execution of workflows on KNIME Server via REST API



Automation: Call Local Workflow

ETL

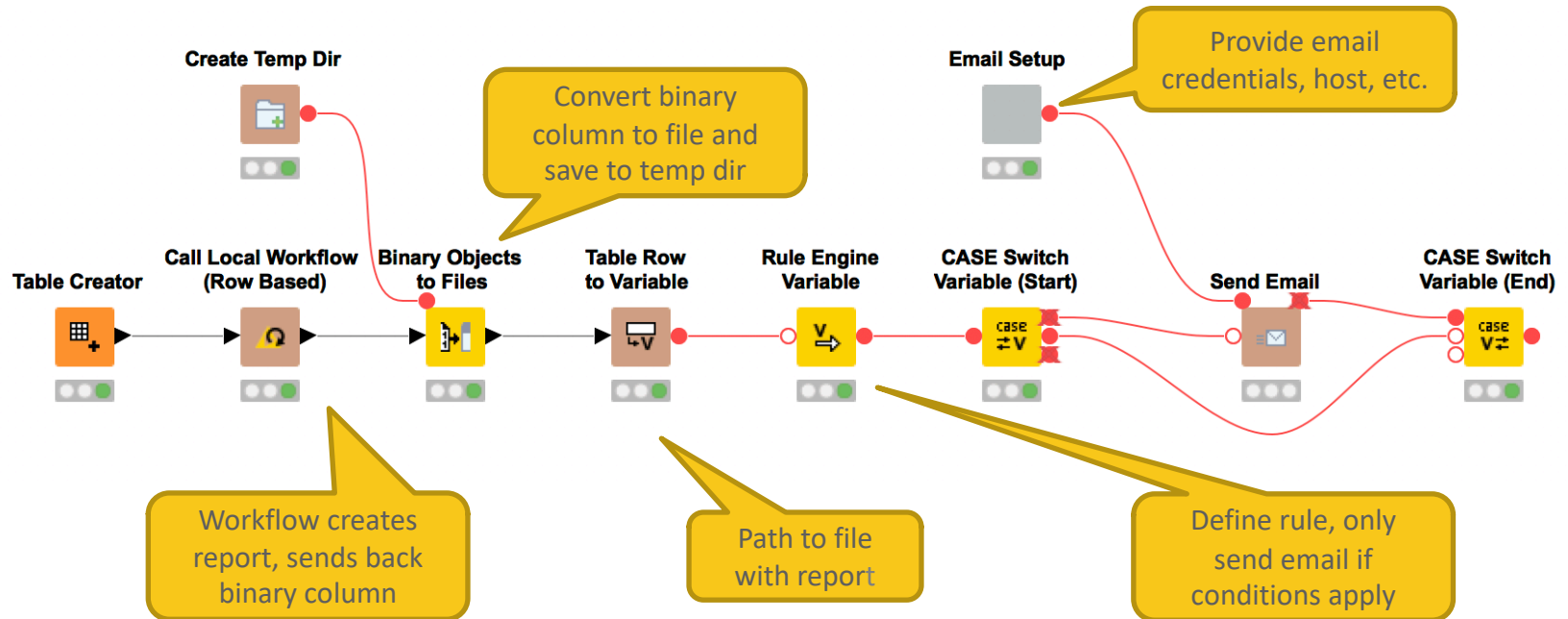
Prediction



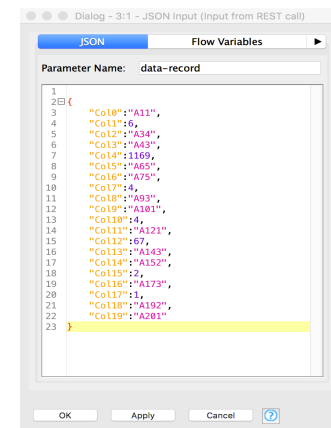
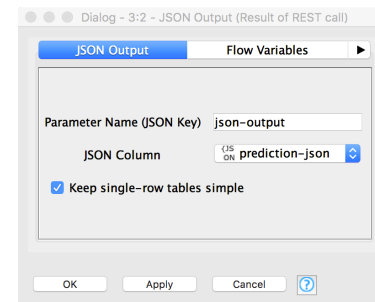
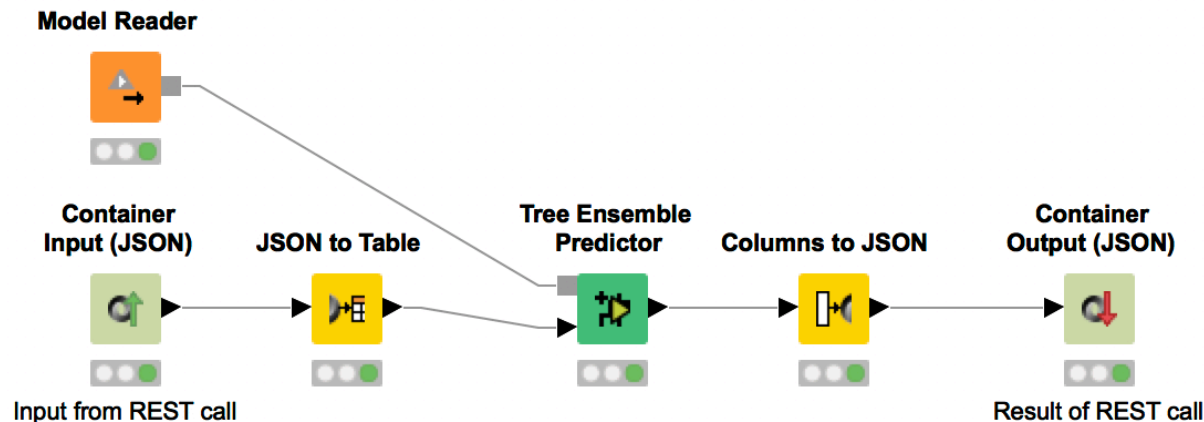
Use Call Local Workflow to Send Conditional Emails with Report

Sometimes, report should be sent under specific circumstances

- E.g. if some KPI is below threshold



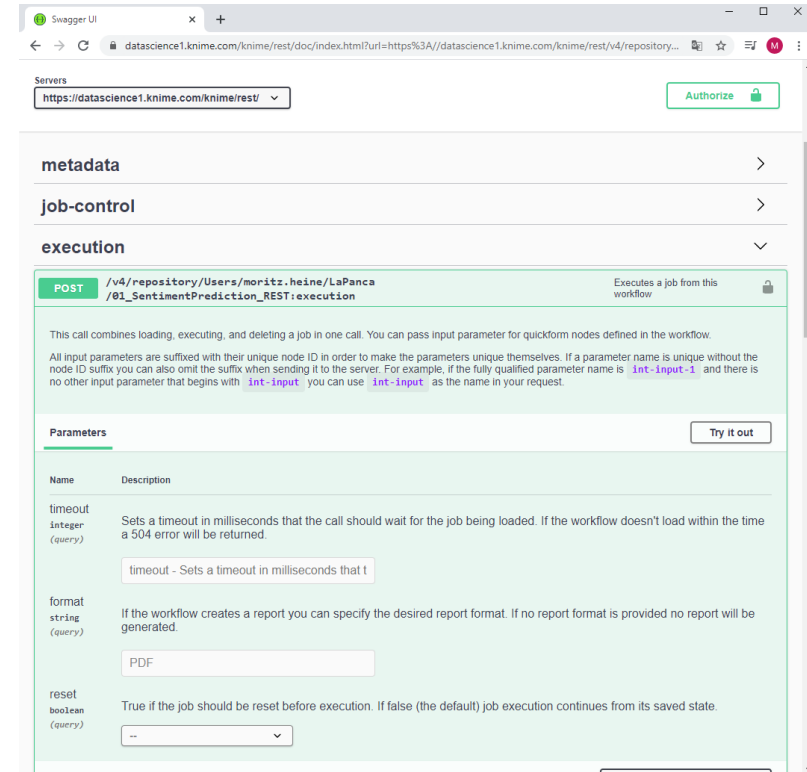
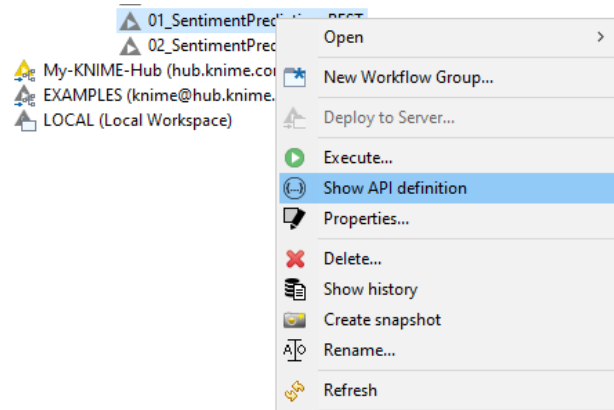
KNIME Server as a REST Resource



<https://www.knime.org/blog/giving-the-knime-server-a-rest>

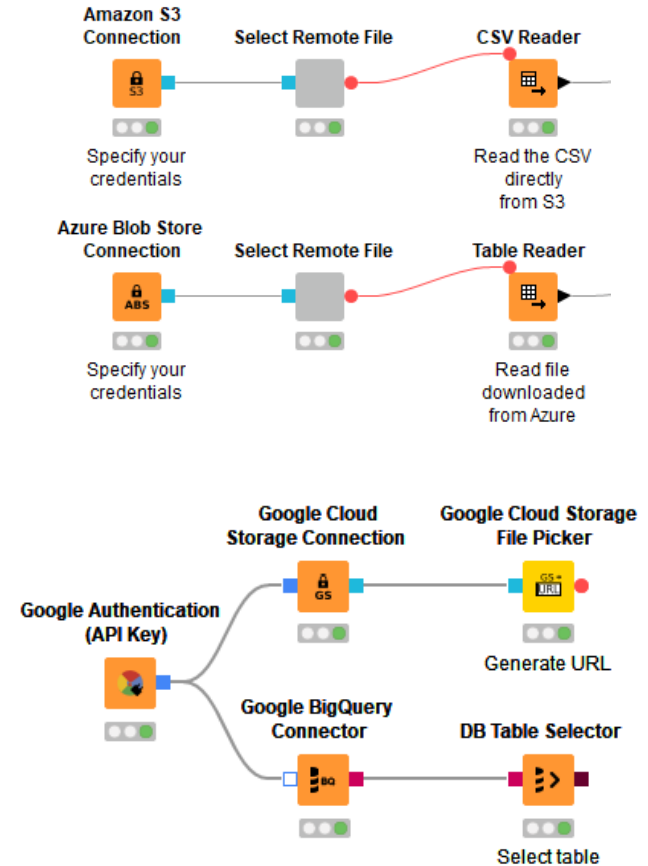
KNIME Server as a REST resource

- Use Swagger, SOAPUI or Chrome extension Postman to explore the HTTP requests and test them



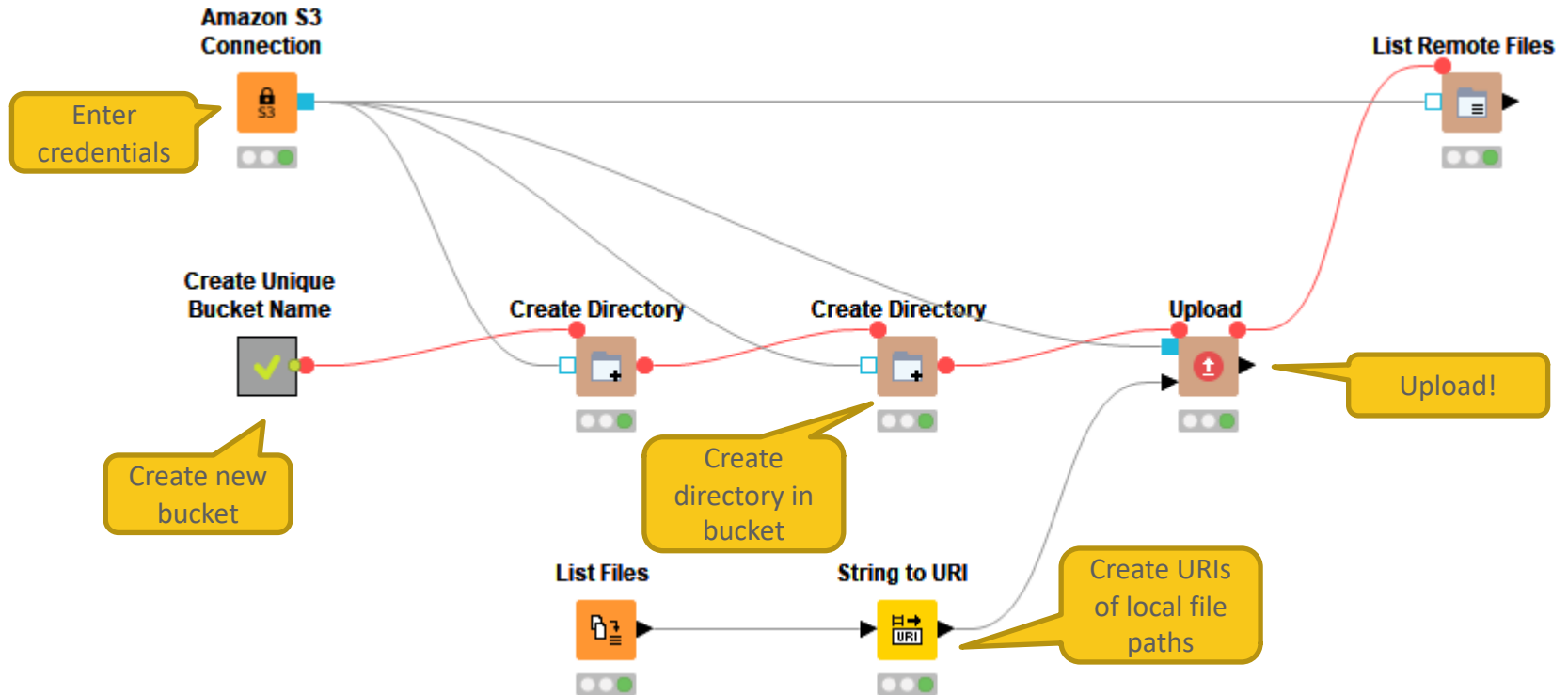
Remote File Handling – Cloud Storage

- Integrate remote data sources from Amazon AWS, Microsoft Azure, and Google Cloud
 - Upload files
 - Download files, or read their content directly into KNIME
 - List files in remote directories
 - Create directories
 - Delete files / directories



Remote File Handling – Cloud Storage

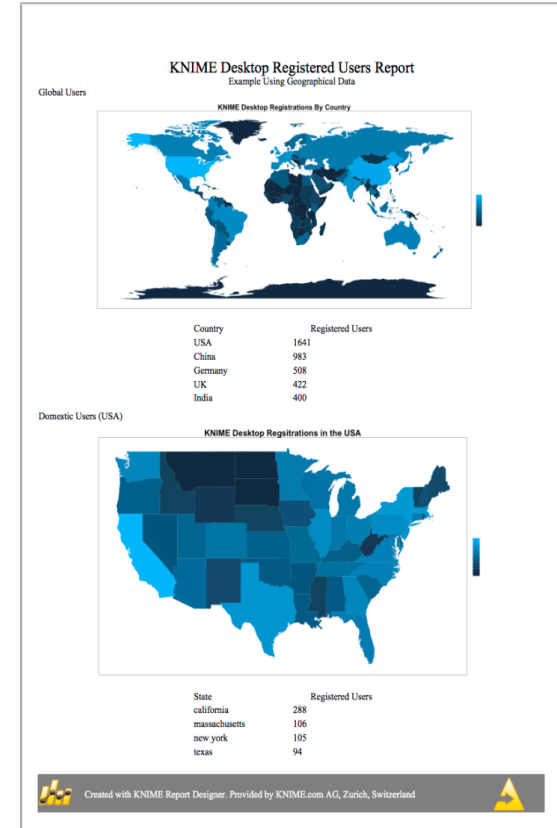
Example: Upload all files from a local directory to Amazon S3






Reporting in KNIME

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, ...)

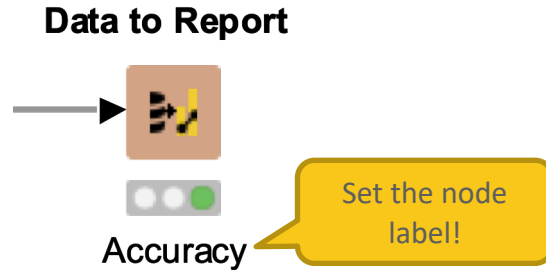


Installation

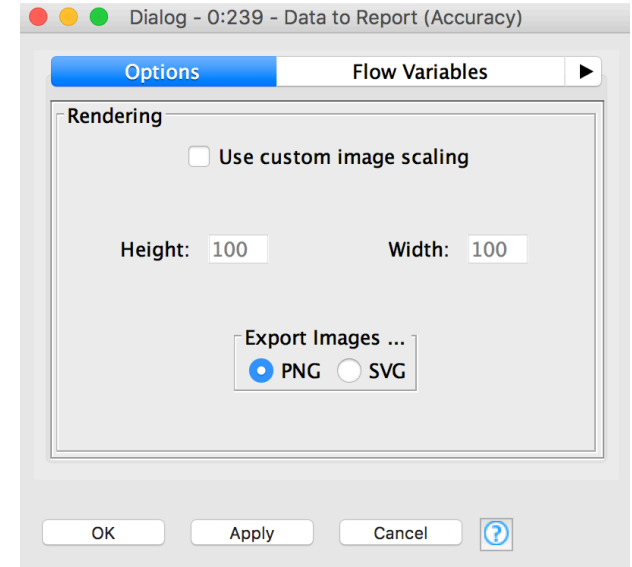
- Can be installed via KNIME -> Install KNIME Extension
- Install the KNIME Report Designer
 - ▼  KNIME Report Designer
 - ▶  BIRT Framework
 - ▶  KNIME Reporting Runtime

New Node: Data to Report

Send a data table to BIRT



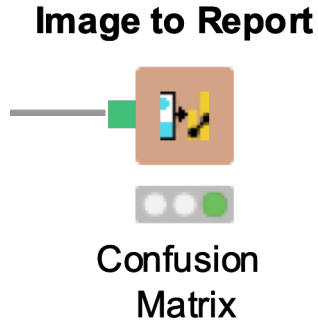
Hint: The node label will be used to identify the data source in the reporting view -> Make sure to use understandable labels if you have more than one data source



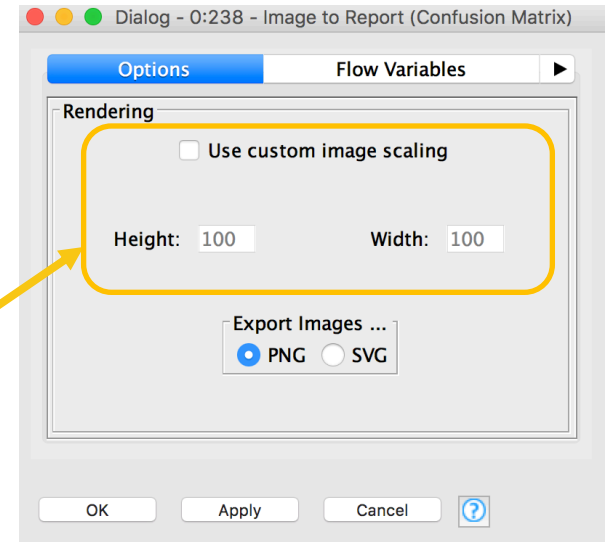
New Node: Image to Report

Send an image to BIRT

- PNG and SVG are supported formats (see node description for details)



Hint: Customize the image size in the Data to Report node to fit the report



Edit the Report

Open the workflow and 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, PMML Writer, Table Writer, 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 Hub Search

Search workflows, nodes, and more...

KNIME Console

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

Reporting Perspective

The screenshot shows the KNIME Reporting Perspective interface. The main window displays a report titled "My Data Mining Report". The interface includes a left sidebar with a tree view of data sources and a palette of report items. The bottom of the window features a "View tabs" section with tabs for "Layout", "Master Page", "Script", and "XML Source". The "Layout" tab is active, showing a report structure with sections for "Overall Model Accuracy", "Confusion Matrix", and "Footer Row".

Click button to create report

Data from KNIME - names of data sources are taken from node label

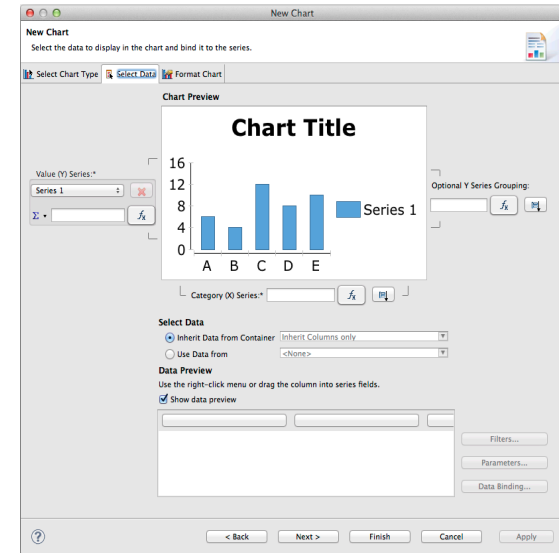
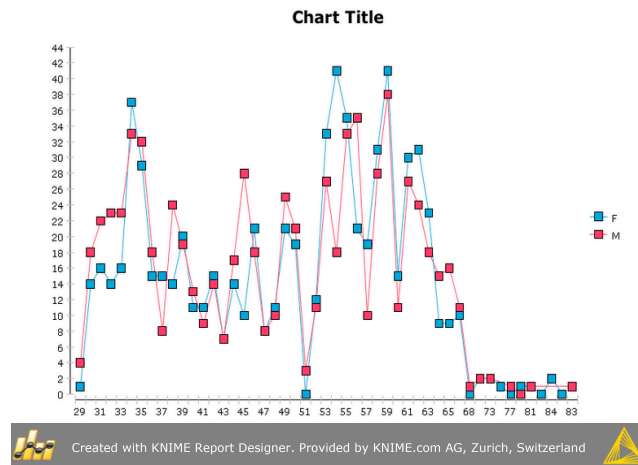
Add report items via drag and drop

Report layout – only structure, data is filled in when creating the report

View tabs

Charting in BIRT

- Many chart types
- Fine control of plot appearance
- Familiar 'Excel Like' interface
- Supports interactivity



Tips & Tricks

- Use an underlying grid to structure the report
- Names of columns should not change
- Use the grouping function to combine results
- Use the Master Layout Tab (For footers etc.)

Exporting Data Exercise

Start with exercise: *Exporting Data*

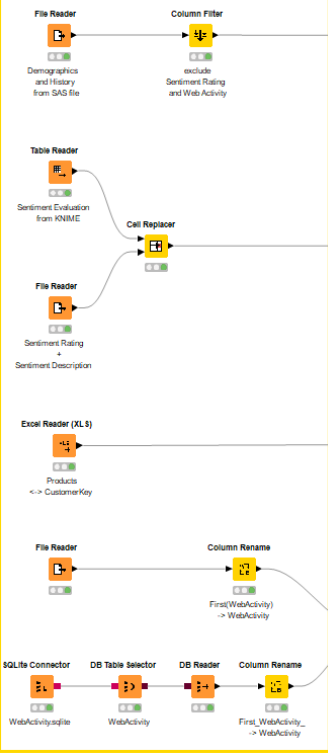
- Write the predictions to a KNIME table
- Write the decision tree model to a PMML model file
- Create a heatmap of the normalized confusion matrix of your model and send it to a BIRT report
- Send your model accuracy to a BIRT report
- Create a simple report showing the overall accuracy and the heatmap of the confusion matrix
- Generate a PDF of your report

Today's Example

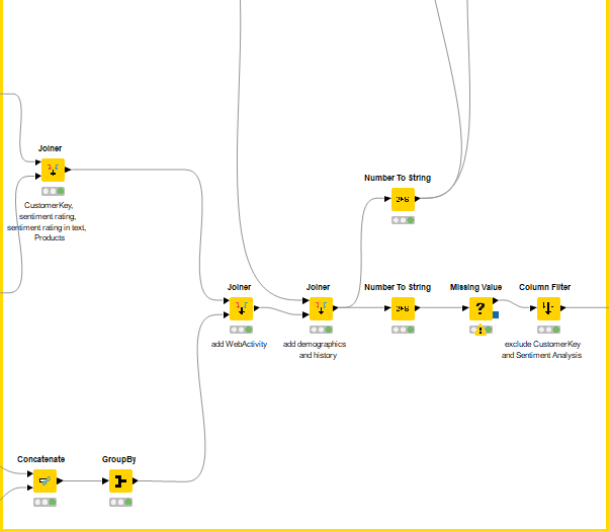
Final Workflow from the KNIME User Training

... and putting all those parts together, you get this final workflow.

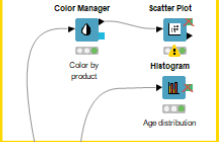
Data Reading



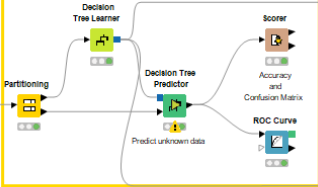
Data Manipulation and Aggregation



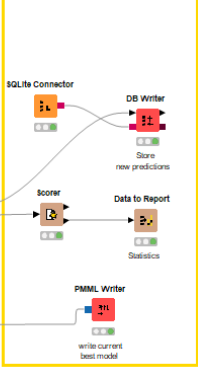
Visualization



Training Predictive Models



Data Export and Reporting





Open for Innovation[®]
KNIME

Thank You!

education@knime.com