

# Overview KNIME



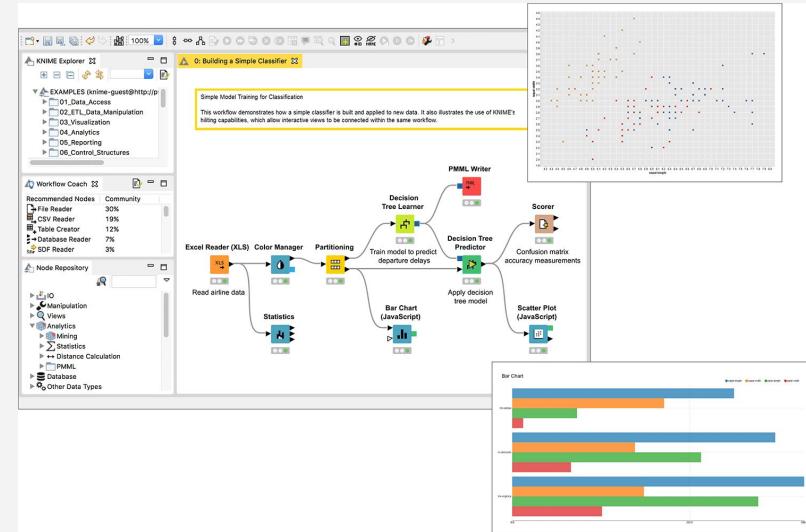
# **Overview**

# **KNIME Analytics**

# **Platform**

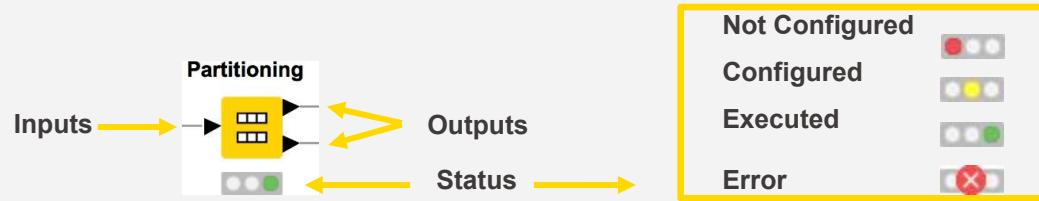
# What is KNIME Analytics Platform?

- 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.

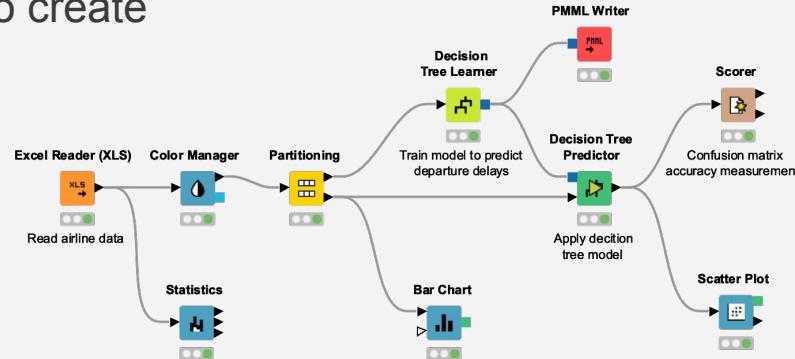


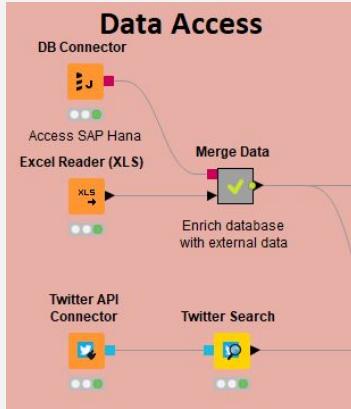
# Visual KNIME Workflows

NODES perform tasks on data



Nodes are combined to create  
**WORKFLOWS**





## ■ Databases

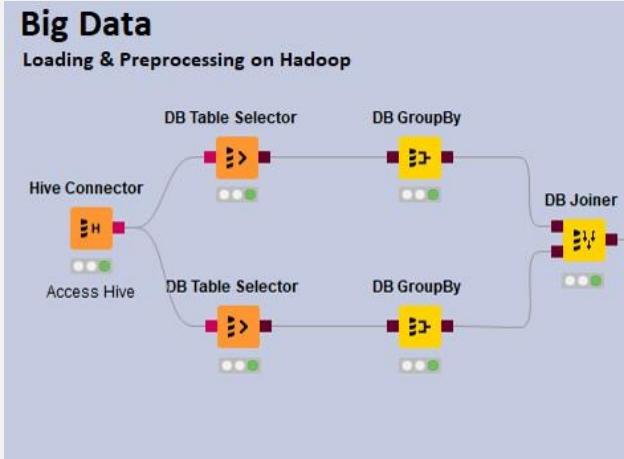
- MySQL, PostgreSQL, Oracle
- Theobald
- any JDBC (DB2, MS SQL Server)
- Amazon DynamoDB

## ■ Files

- CSV, txt, Excel, Word, PDF
- SAS, SPSS
- XML, JSON, PMML
- Images, texts, networks

## ■ Other

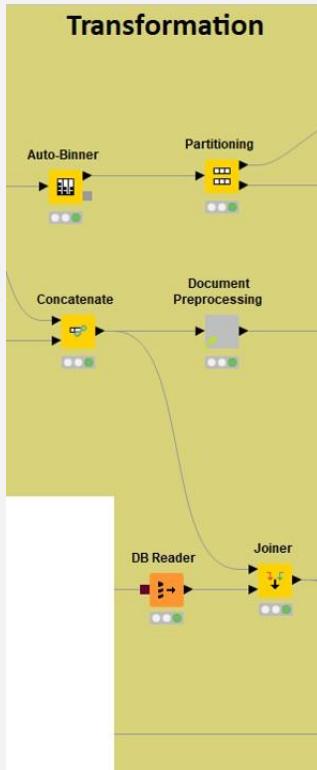
- Twitter, Google
- Amazon S3, Azure Blob Store
- Sharepoint, Salesforce
- Kafka
- REST, Web services



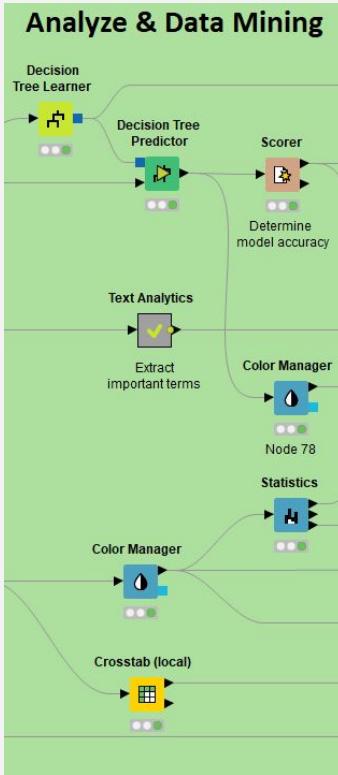
- Spark & Databricks
- HDFS support
- Hive
- Impala
- In-database processing



# Transformation

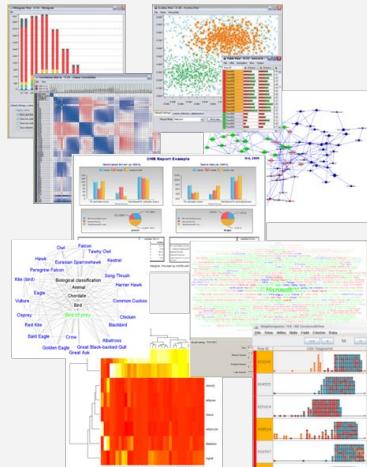
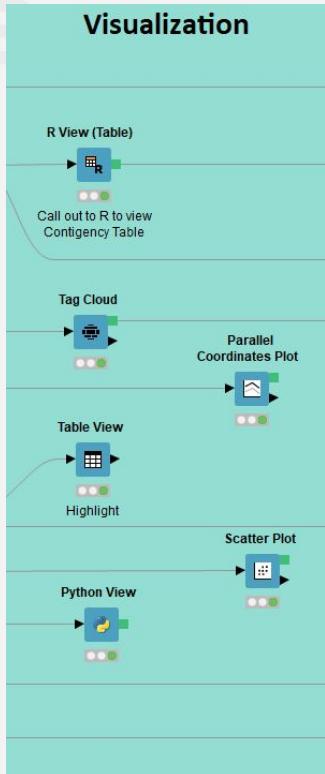


- Preprocessing
  - Row, column, matrix based
- Data blending
  - Join, concatenate, append
- Aggregation
  - Grouping, pivoting, binning
- Feature Creation and Selection



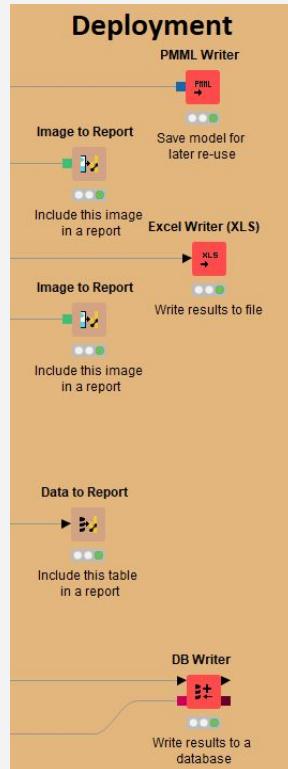
- Regression
  - Linear, logistic
- Classification
  - Decision tree, ensembles, SVM, MLP, Naïve Bayes
- Clustering
  - k-means, DBSCAN, hierarchical
- Validation
  - Cross-validation, scoring, ROC
- Deep Learning
  - Keras, DL4J
- External
  - R, Python, Weka, H2O, Keras

# Visualization



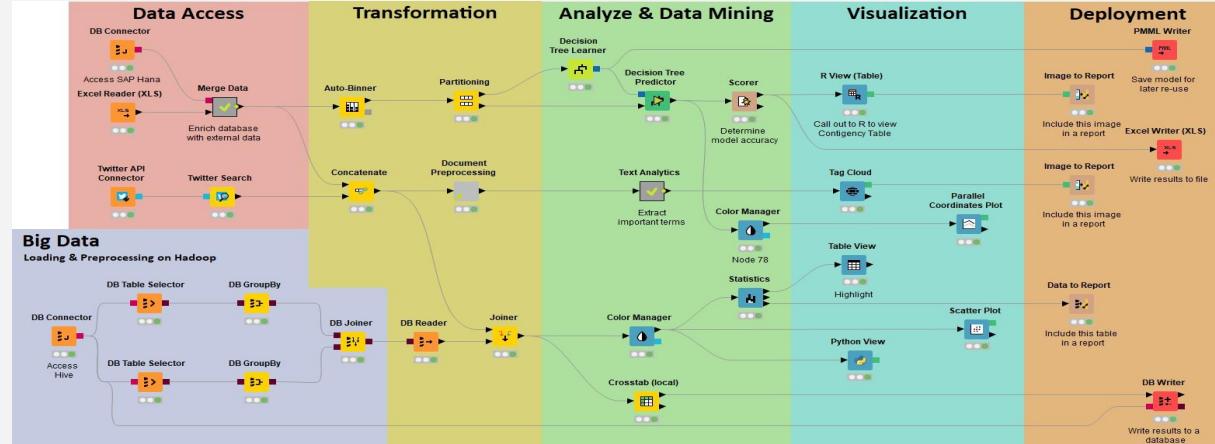
- Interactive Visualizations
- JavaScript-based nodes
  - Scatter Plot, Box Plot, Line Plot
  - Networks, ROC Curve, Decision Tree
  - Plotly Integration
  - Adding more with each release!
- Misc
  - Tag cloud, open street map, molecules
- Script-based visualizations
  - R, Python

# Deployment



- Database
- Files
  - Excel, CSV, txt
  - XML
  - PMML
  - to: local, KNIME Server, Amazon S3, Azure Blob Store
- 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 /

## Transformatio

n  
Row  
Column  
n  
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

## Visualiza

R  
JFreeChart  
JavaScript  
Plotly  
Community / 3rd

## Deploym

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

# Install KNIME Analytics Platform

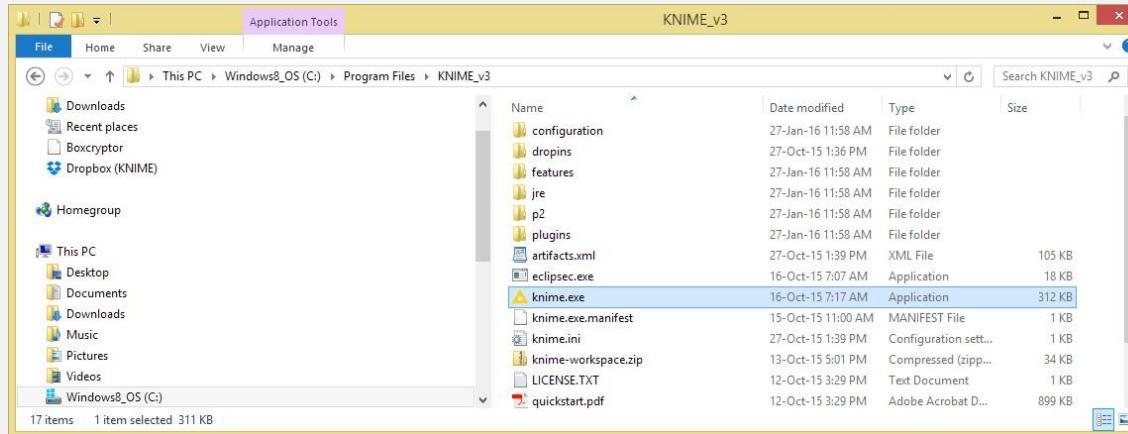
- Select the KNIME version for your computer:
  - Mac
  - Windows – 32 or 64 bit
  - Linux
- Download archive and extract the file, or download installer package and run it

Windows	
KNIME Analytics Platform for Windows (installer) <i>The installer adds an icon to the desktop and suggests suitable memory settings</i>	<a href="#">32 Bit (393.38 MB)</a> <a href="#">64 Bit (396.38 MB)</a>
KNIME Analytics Platform for Windows (self-extracting archive) <i>The self-extracting archive only creates a folder holding the KNIME installation</i>	<a href="#">32 Bit (396.87 MB)</a> <a href="#">64 Bit (400.72 MB)</a>
KNIME Analytics Platform for Windows (zip archive)	<a href="#">32 Bit (466.11 MB)</a> <a href="#">64 Bit (470.07 MB)</a>
Linux	
KNIME Analytics Platform for Linux	<a href="#">64 Bit (417.21 MB)</a>
Mac	
KNIME Analytics Platform for Mac OSX (10.11 and above)	<a href="#">64 Bit (388.44 MB)</a>

# Start KNIME Analytics Platform

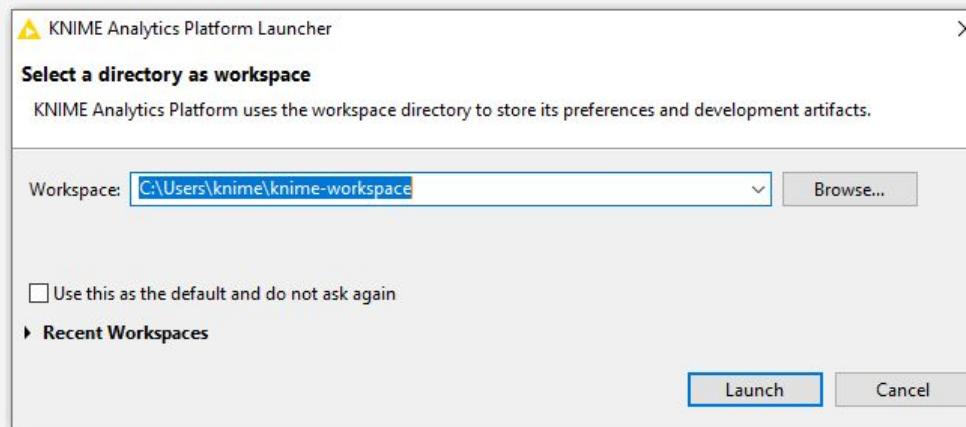


- Use the shortcut created by the installer
- Or go to the installation directory and launch KNIME via the knime.exe



# 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 Analytics Platform Workbench

The screenshot shows the KNIME Analytics Platform Workbench interface. The central workspace displays a workflow titled "My first Workflow". The workflow consists of the following nodes and connections:

- File Reader** (orange) → **Row Filter** (yellow)
- Row Filter** (yellow) → **Column Filter** (yellow)
- Column Filter** (yellow) → **Table Writer** (red)

The **Row Filter** node has the configuration: "keep only records born in the US". The **Column Filter** node has the configuration: "remove gender". The **Table Writer** node has the configuration: "Write table".

The left sidebar includes:

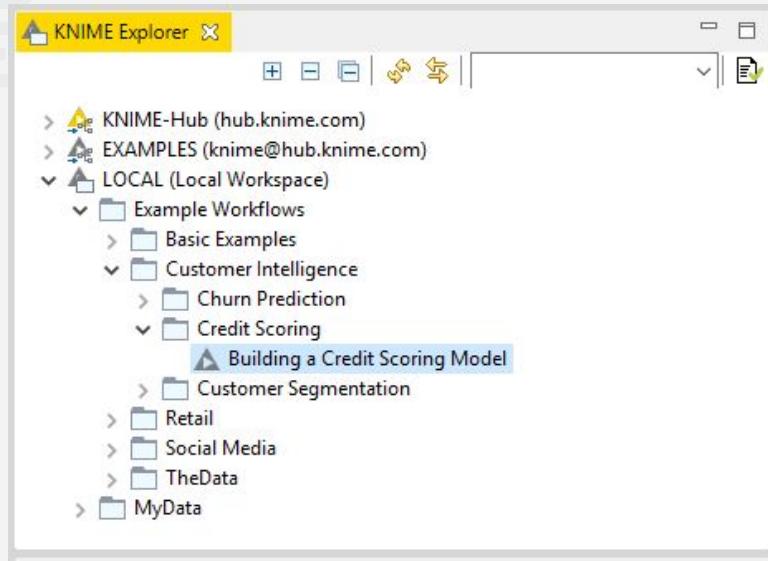
- KNIME Explorer**: Shows sections like "My-KNIME-Hub", "EXAMPLES", and "LOCAL (Local Workspace)" with "Basic Examples" highlighted.
- Workflow Coach**: Lists recommended nodes by community usage.
- Node Repository**: Lists categories like "IO", "Manipulation", "Views", "Analytics", "DB", etc.

The right sidebar includes:

- Description**: A detailed description of the **Row Filter** node, explaining its purpose and configuration options.
- KNIME Hub Search**: A search bar for workflows, nodes, and more.

The bottom right corner shows a preview of the data table output from the workflow, which includes columns such as ID, age, workclass, fnlwgt, education, education-num, marital-status, occupation, and relationship.

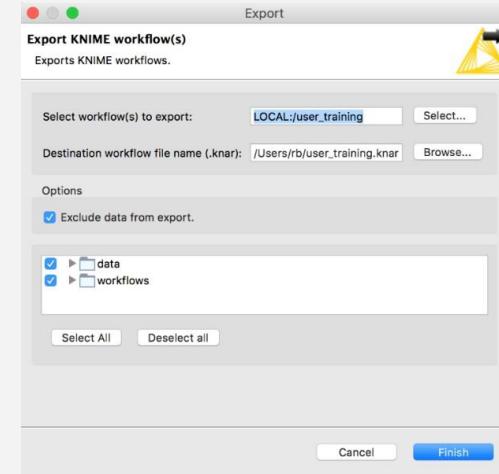
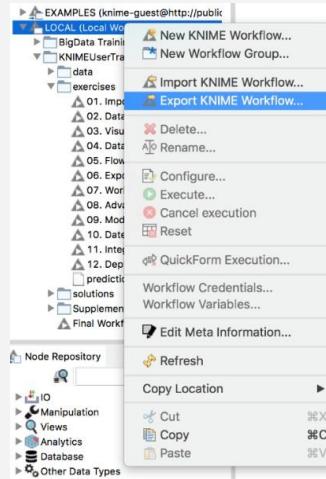
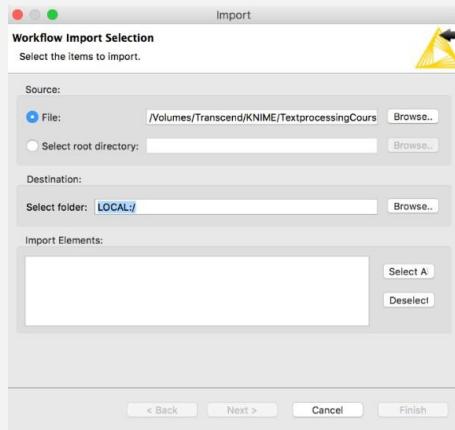
# KNIME Explorer



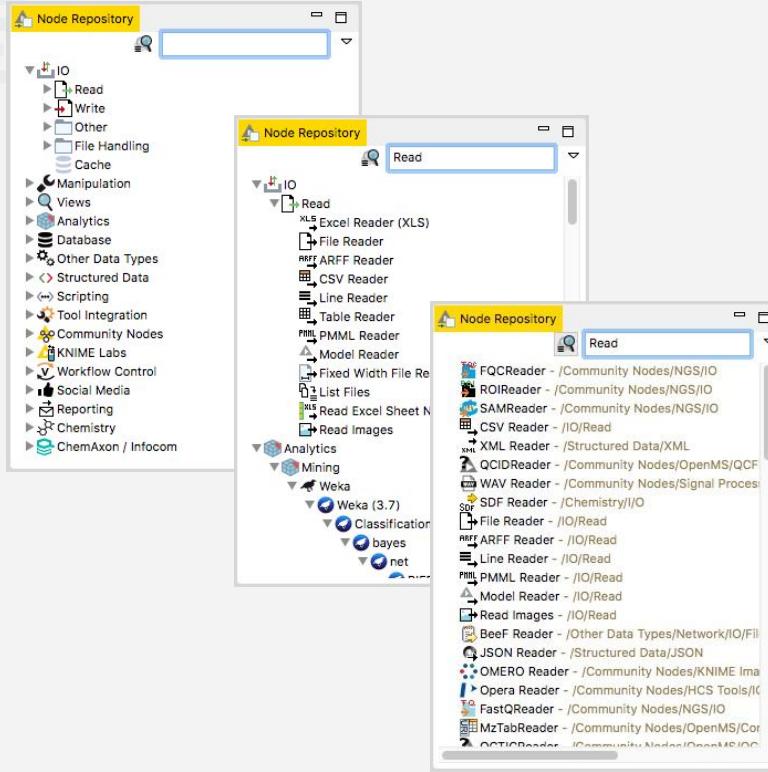
- In LOCAL you can access your own workflow projects.
- The Explorer toolbar on the top has a search box and buttons to
  - ➡ select the workflow displayed in the active editor
  - ➡ refresh the view
- The KNIME Explorer can contain 4 types of content:
  - Workflows
  - Workflow groups
  - Data files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

# Creating New Workflows, Importing and Exporting

- Right-click inside the KNIME Explorer to create a new workflow or a workflow group, or to import a workflow
- Right-click the workflow or workflow group to export

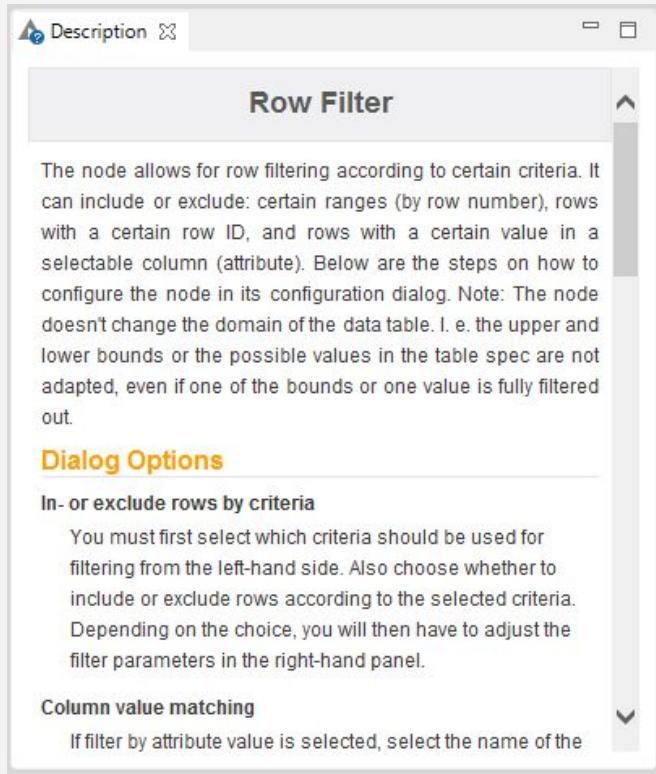


# Node Repository



- 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

# Description



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

# Workflow Description

The screenshot shows the 'Description' window for a workflow named 'My\_First\_Workflow'. The window has a yellow header bar with a question mark icon and a close button. Below the header, the workflow title is displayed as 'My\_First\_Workflow' with a pencil icon for editing. The 'Title' field contains 'My First Workflow'. The 'Description' section states: 'This workflow reads data, removes uninteresting columns and rows, and writes the resulting data table to a CSV file.' The 'Tags' section lists 'Example Workflow', 'CSV', and 'Data Manipulation'. The 'Links' section provides links to the 'KNIME Homepage', 'KNIME Hub', and 'KNIME Forum'. The 'Creation Date' is listed as '2019-7-2'. The 'Author' is 'Ana Vedoveli'.

Description

My\_First\_Workflow

Title My First Workflow

Description

This workflow reads data, removes uninteresting columns and rows, and writes the resulting data table to a CSV file.

Tags

Example Workflow CSV Data Manipulation

Links

- KNIME Homepage
- KNIME Hub
- KNIME Forum

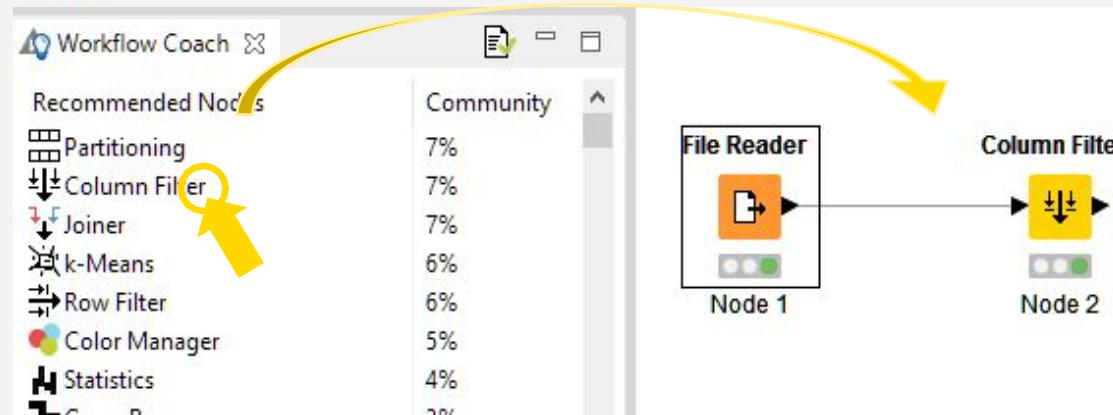
Creation Date 2019-7-2

Author Ana Vedoveli

- 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



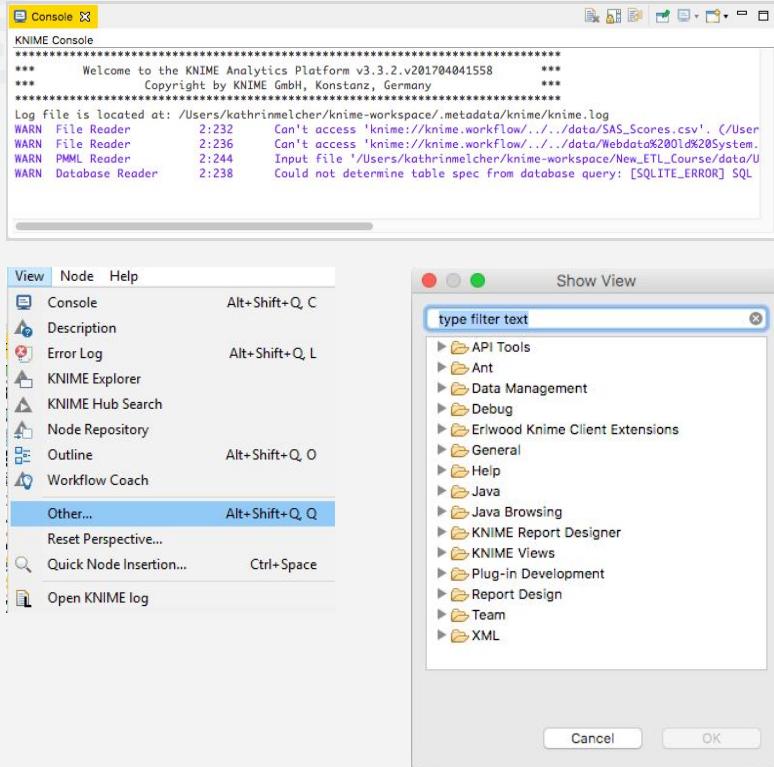
# Node Monitor

- By default the Node Monitor shows you the output table of the node selected in the workflow editor
- Click on the three dots on the upper right to show the flow variables, configuration, etc.

The screenshot shows the Node Monitor window with the following details:

- Console Node Monitor**: The title bar.
- Node:** Get Customers from Database (0:1207)
- State:** EXECUTED
- Port Output:** Port 0
- Load data**: A button to refresh the data.
- Table Data:** A grid of customer information. The columns are: ID, MaritalStatus, Gender, EstimatedYearlyIncome, NumberOfContracts, Age, Available401K, CustomerV, and Products. The data includes rows for various customer IDs with their respective details.
- Context Menu:** A dropdown menu is open on the right side of the window, listing options:
  - Show Output Table (selected)
  - Show Variables
  - Show Configuration
  - Show Entire Configuration
  - Show Node Timing Information
  - Show Graph Annotations

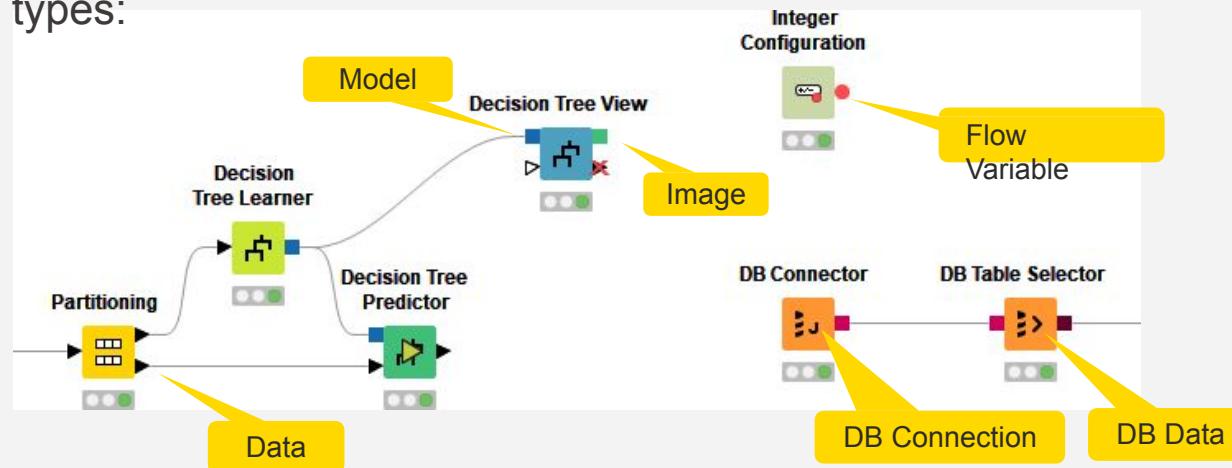
# Console and Other Views



- 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.

# 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:



# More on Nodes...

- A node can have 4 states:



## Not Configured:

The node is waiting for configuration or incoming data.



## Configured:

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



## Executed:

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

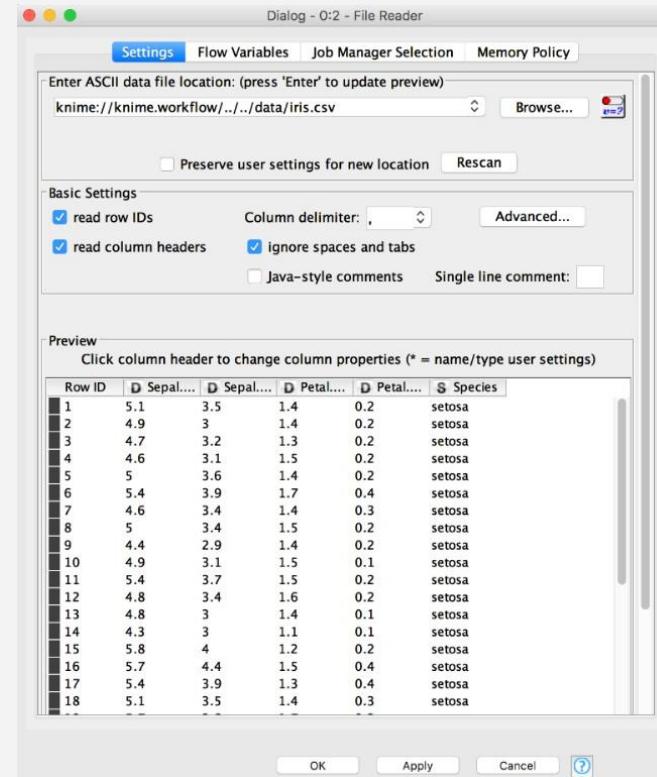


## Error:

The node has encountered an error during execution.

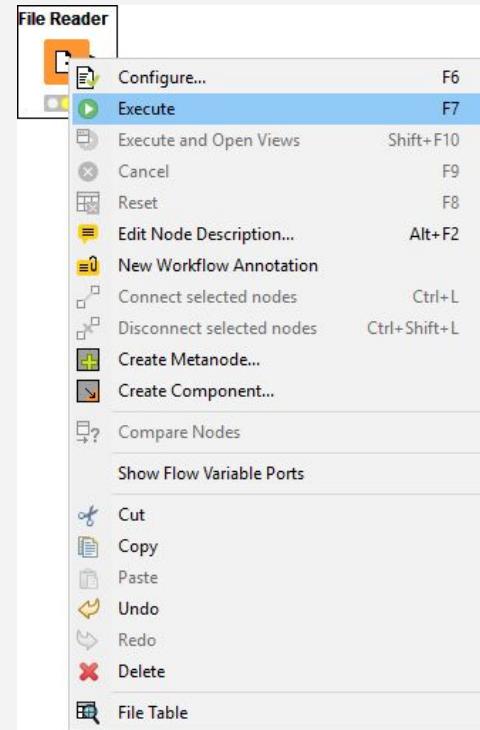
# Node Configuration

- 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



# Tool Bar

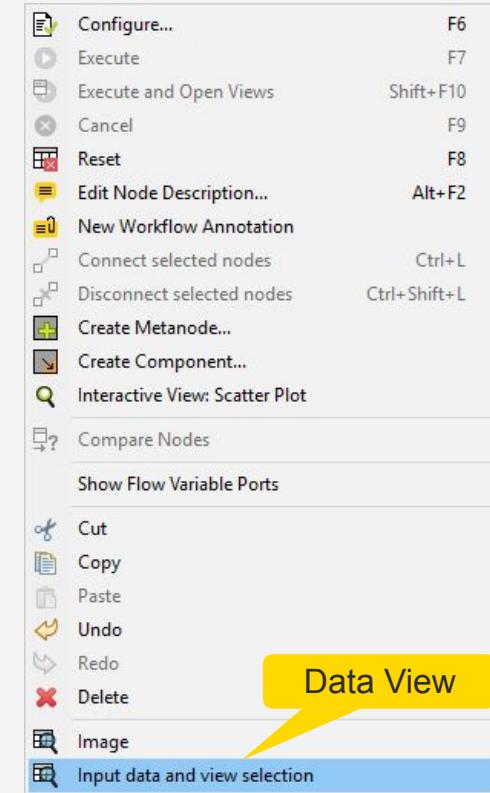
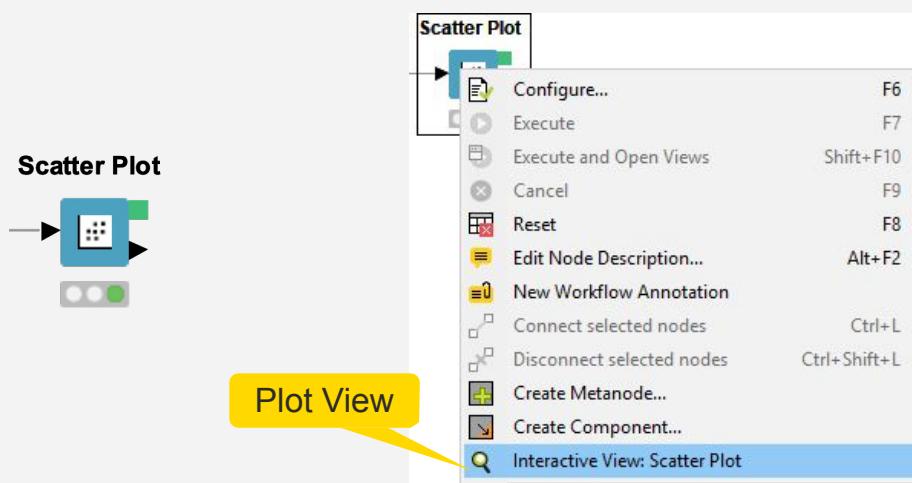


The buttons in the toolbar can be used for the active workflow. The most important buttons are:

- ▶ 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

# Node Views

- Right-click node to inspect the execution results by
  - selecting output ports (last option in the context menu) to inspect tables, images, etc.
  - selecting Interactive View to open visualization results in a browser



# 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



# Getting Started: KNIME Hub

■ Place to search and share

- Workflows
- Nodes
- Components
- Extensions

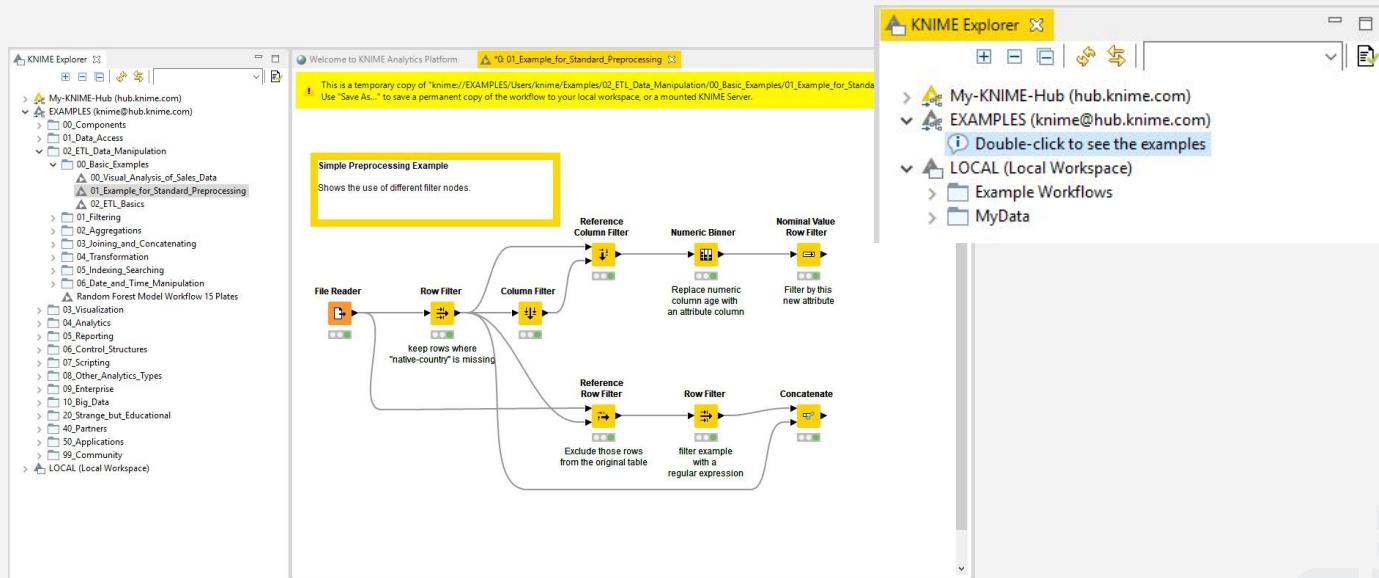
The screenshot shows the KNIME Hub search interface. At the top, there's a search bar with the query "Sentiment Analysis". Below it, a large number "350 results" is displayed. Underneath, there are four workflow cards, each with a thumbnail, title, category tags (e.g., "deep learning", "keras", "text classification"), and a brief description. The first two workflows are identical, showing how to train a simple neural network for text classification using Keras and LSTM. The third workflow is similar but with a different file name. The fourth workflow is titled "Sentiment Analysis (Classification) of Documents" and describes importing text from a CSV file, converting it to documents, preprocessing them, and transforming them.

The screenshot shows the main homepage of the KNIME Hub. It features a header with the KNIME logo and a "Sign in" button. Below the header, a welcome message reads: "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." A search bar is present. On the right side, there are four large numerical statistics: 993 Nodes, 265 Components, 2 541 Workflows, and 211 Extensions. Below these stats are three promotional cards: one for "Getting started" with a "How to" guide and a diagram; another for "Courses, Publications, Authors and Institutions"; and a third for the "Forum" with the text "Get help from our community and help others".

<https://hub.knime.com>

# Getting Started: KNIME Example Server

- Connect via KNIME Explorer to a public repository with large selection of example workflows for many, many applications



# Hot Keys (for Future Reference)

Task	Hot key	Description
Node Execution	F6	opens the configuration window of the selected node
	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

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

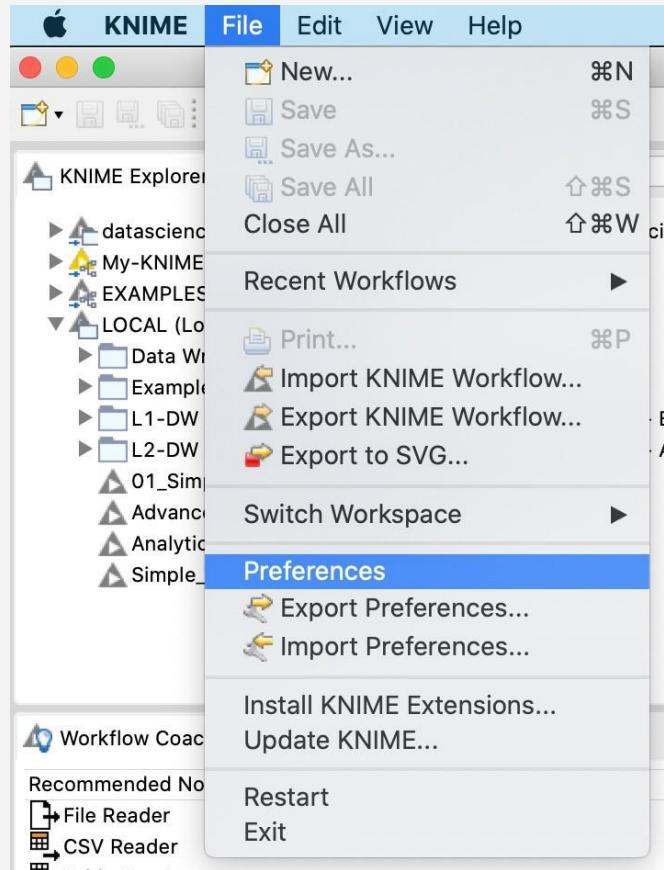
# Adding a Local Update Site

- Download the update site as zip
  - KNIME Updates: <https://www.knime.com/downloads/update>
  - Community Updates: <https://www.knime.com/community>

# Adding a Local Update Site

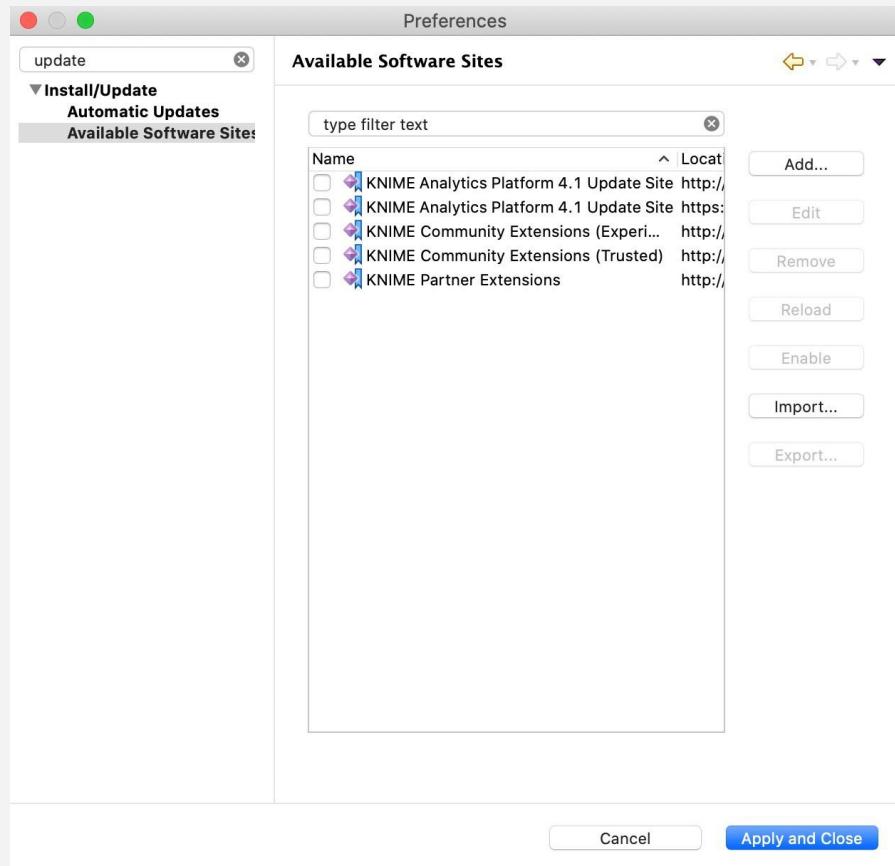
Open KNIME Analytics Platform and go to the preference page by clicking on

File -> Preferences



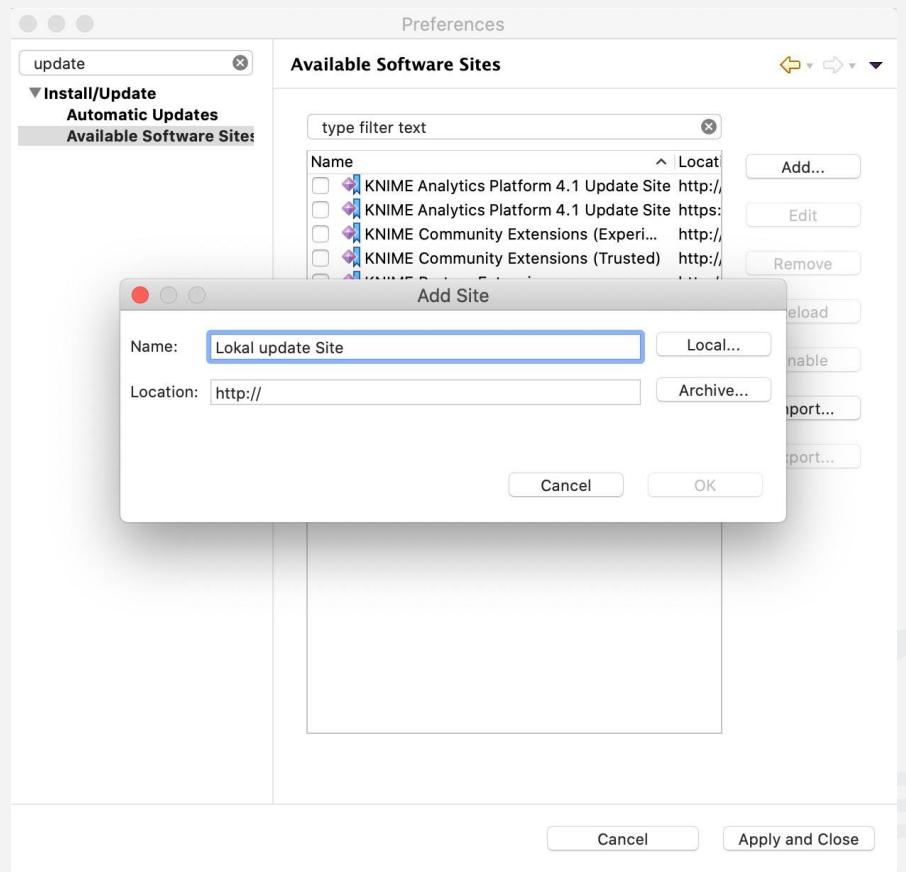
# Adding a Local Update Site

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

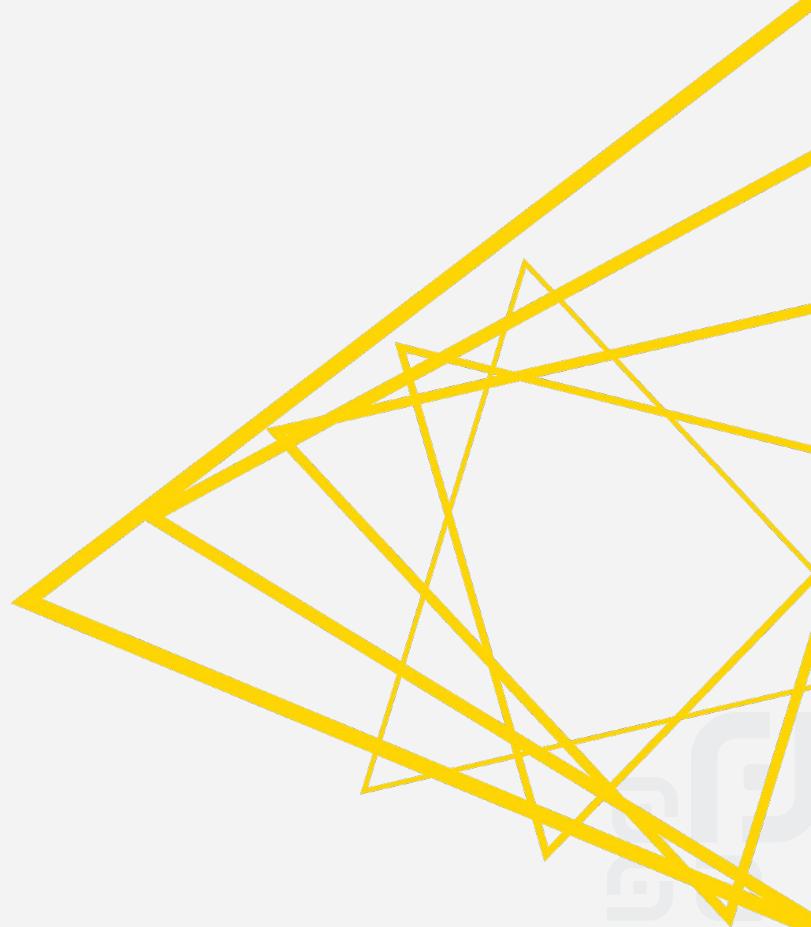


# Adding a Local Update Site

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

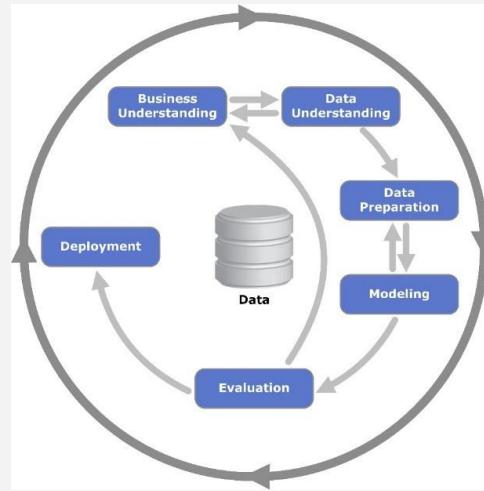


# **Today's Example**

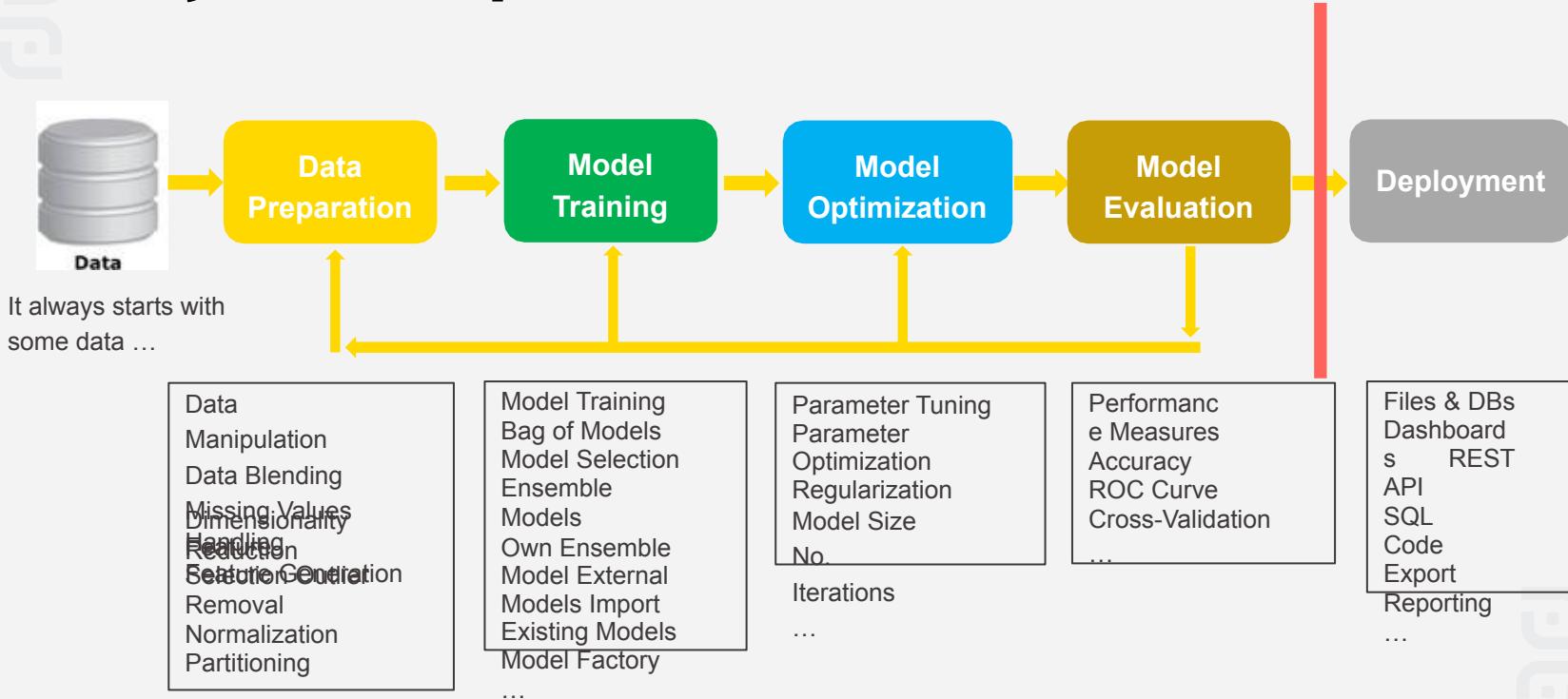


# Today's Example: Churn Prediction

- 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

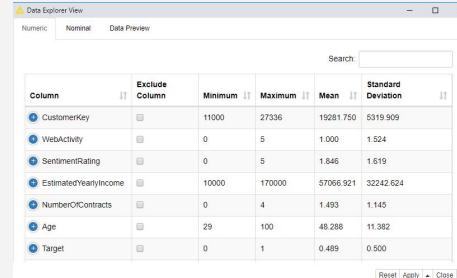


# Today's Example: Churn Prediction



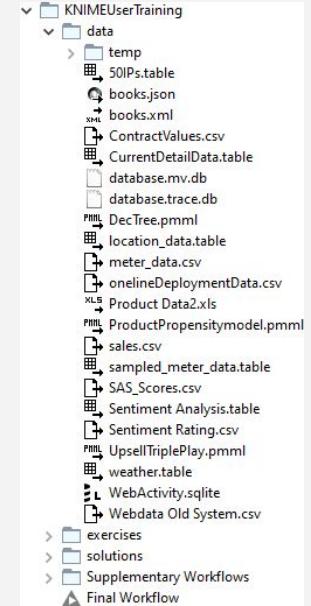
# The Data

- 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

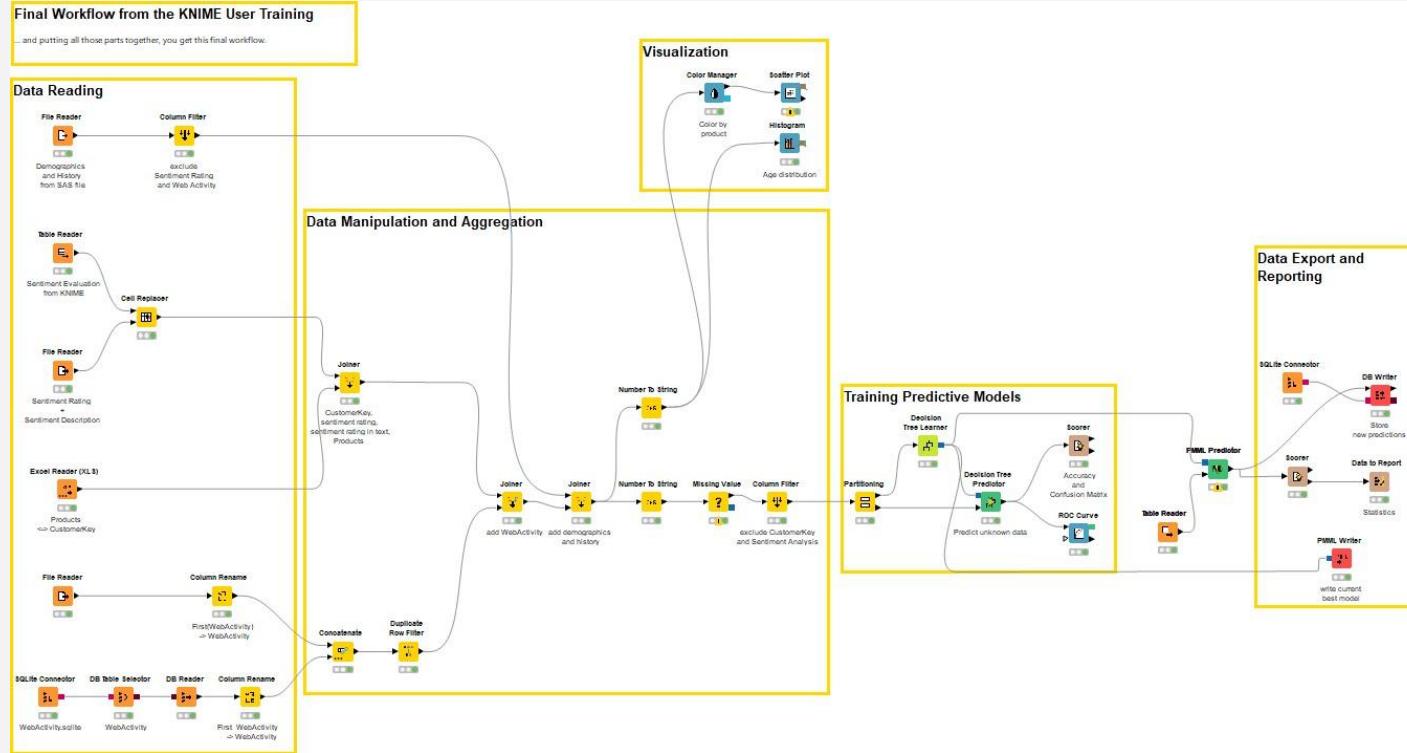


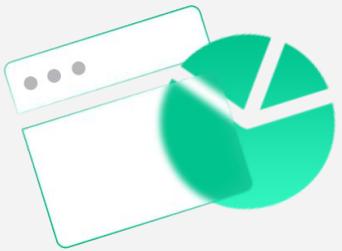
A screenshot of the KNIME Data Explorer View window. It shows a table with columns: Column, Exclude Column, Minimum, Maximum, Mean, and Standard Deviation. The table contains data for various customer attributes:

Column	Exclude Column	Minimum	Maximum	Mean	Standard Deviation
CustomerKey		11000	27356	19281.750	5319.909
WebActivity		0	5	1.000	1.524
SentimentRating		0	5	1.846	1.619
EstimatedYearlyIncome		10000	170000	57066.921	32242.624
NumberOfContracts		0	4	1.493	1.145
Age		29	100	48.288	11.382
Target		0	1	0.489	0.500



# Today's Example: Churn Prediction





# Terima Kasih

SIB Cycle 6 | 2024



[www.greatedu.co.id](http://www.greatedu.co.id)