

Introdução à Plataforma KNIME Analytics

Fatec

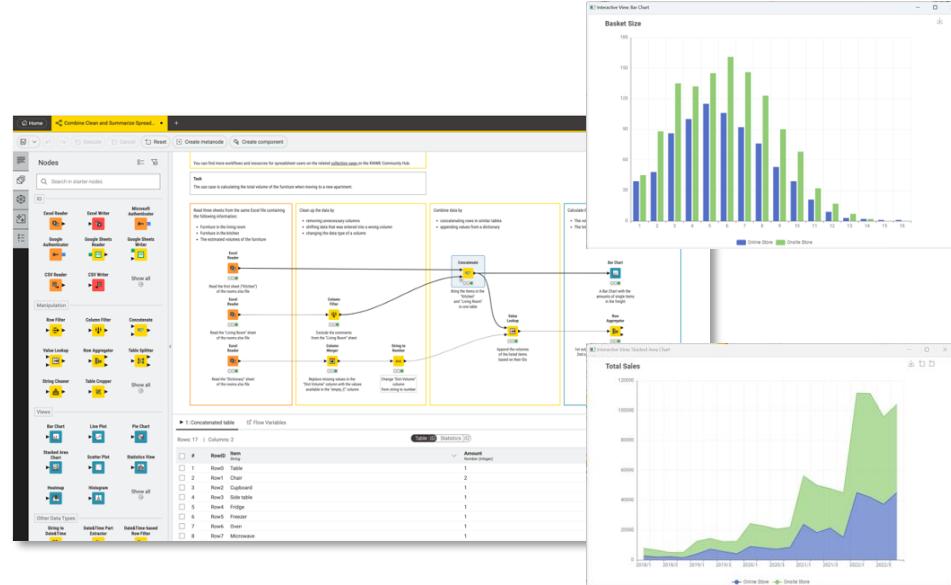
2025

KNIME Analytics Platform

- Ferramenta projetada para coleta, manipulação, análise e modelagem de pipelines de dados por meio de programação visual
- Permite análises preditivas com técnicas de aprendizado de máquina
- Oferece vários exemplos prontos para uso que podem ser facilmente parametrizados
- Suporta a combinação de dados entre arquivos de texto simples, bancos de dados, documentos, imagens, redes e dados baseados no Hadoop em um único fluxo de trabalho visual.

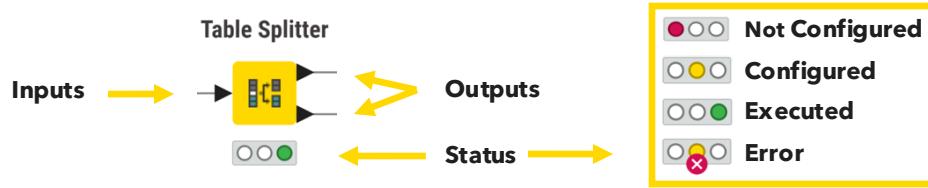
What is KNIME Analytics Platform

- A tool to make sense of your data
 - Data analysis, Data Science, Data Engineering
 - Transformation, visualization, reporting
- Open source
- Visual programming paradigm
 - No coding required

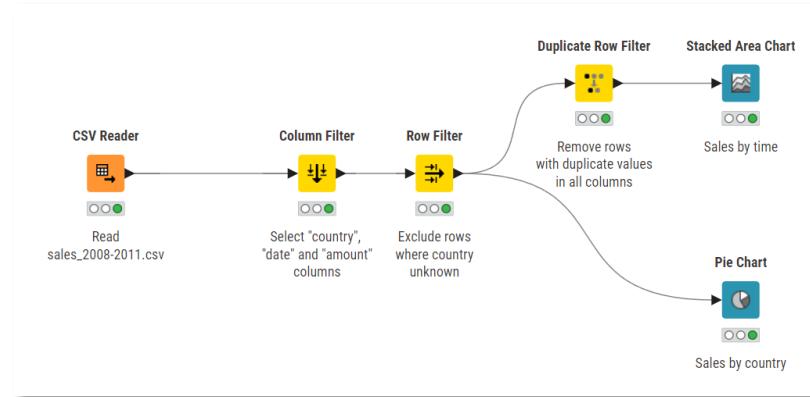


Nodes and Workflows

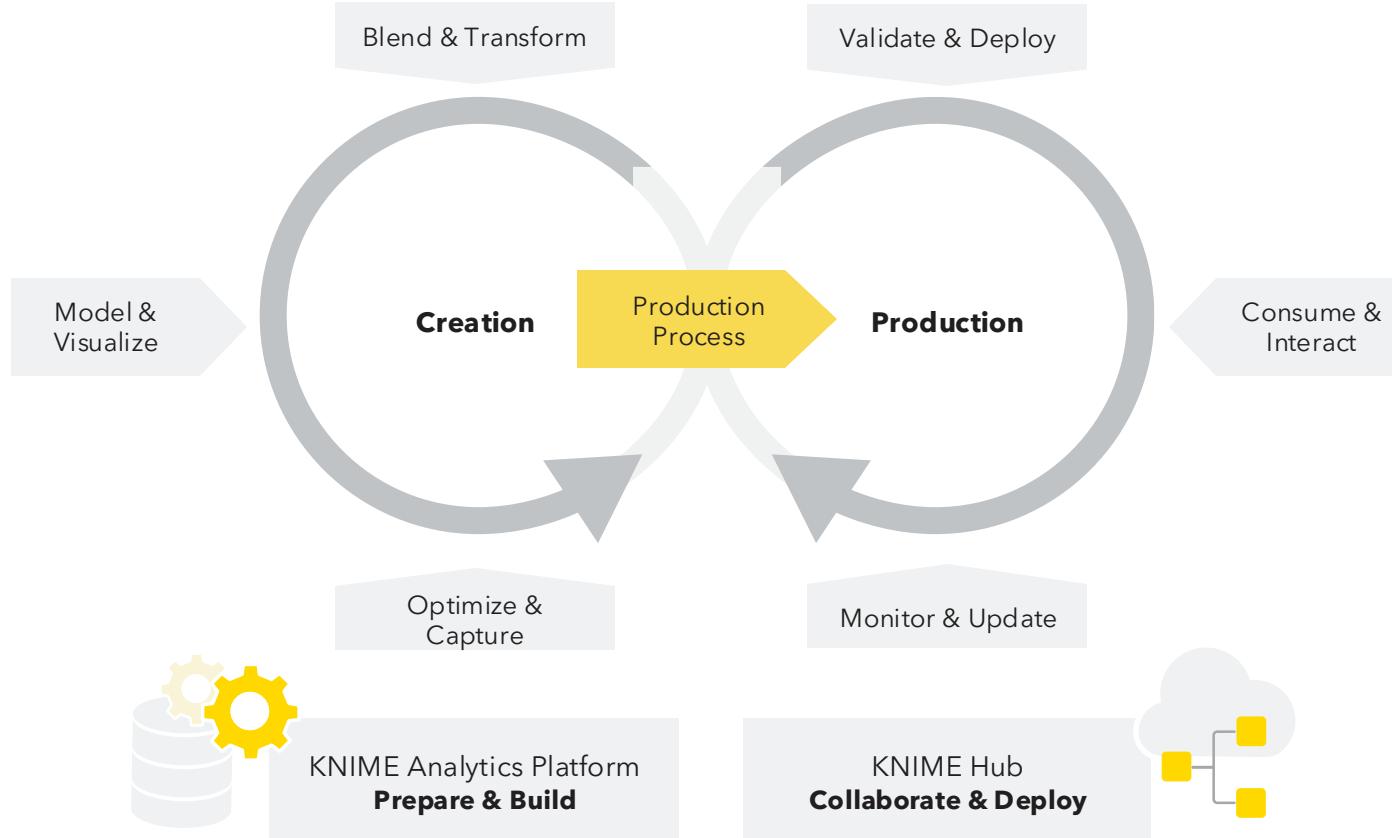
- **NODES** perform tasks on data



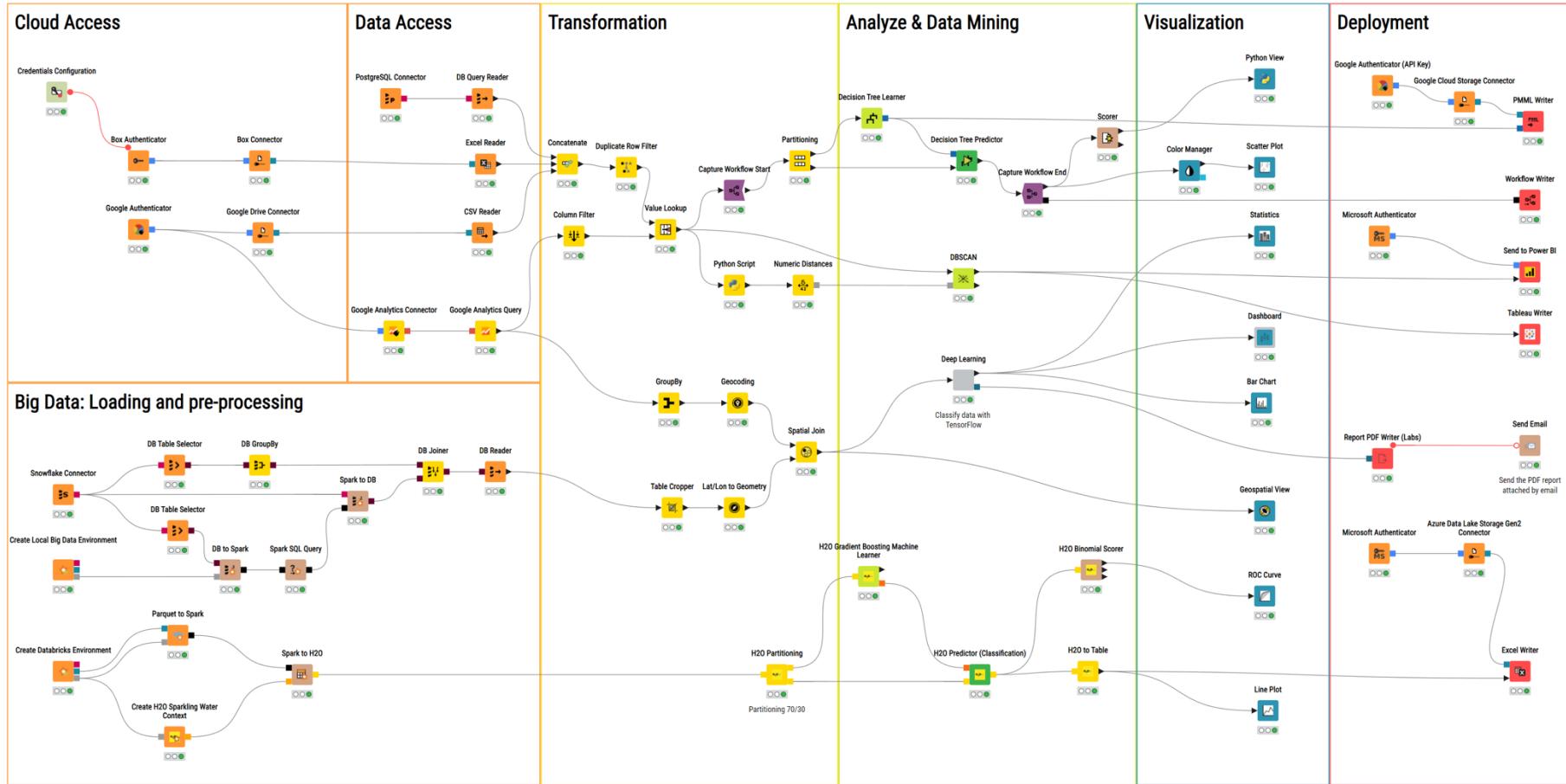
- Nodes are combined to create **WORKFLOWS**



The Data Science Life Cycle



Over 4000 Nodes included



Data Access

Cloud Access

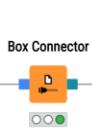
Credentials Configuration



Box Authenticator



Box Connector



Google Authenticator



Google Drive Connector



Data Access

PostgreSQL Connector



DB Query Reader



Excel Reader



CSV Reader



Google Analytics Connector



Google Analytics Query



Files

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

Databases

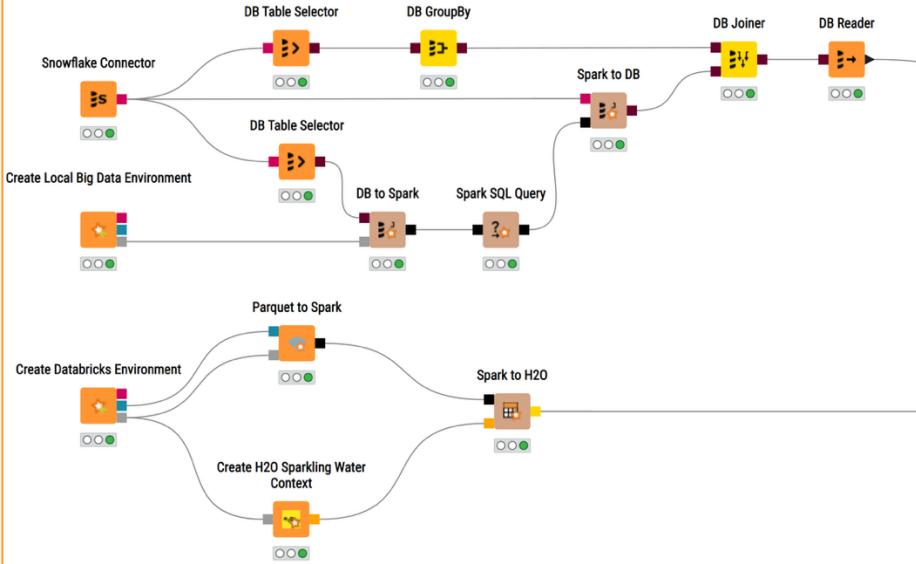
- MySQL, PostgreSQL, Oracle
- Theobald
- any JDBC (DB2, MS SQL Server)
- Amazon DynamoDB

Other

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

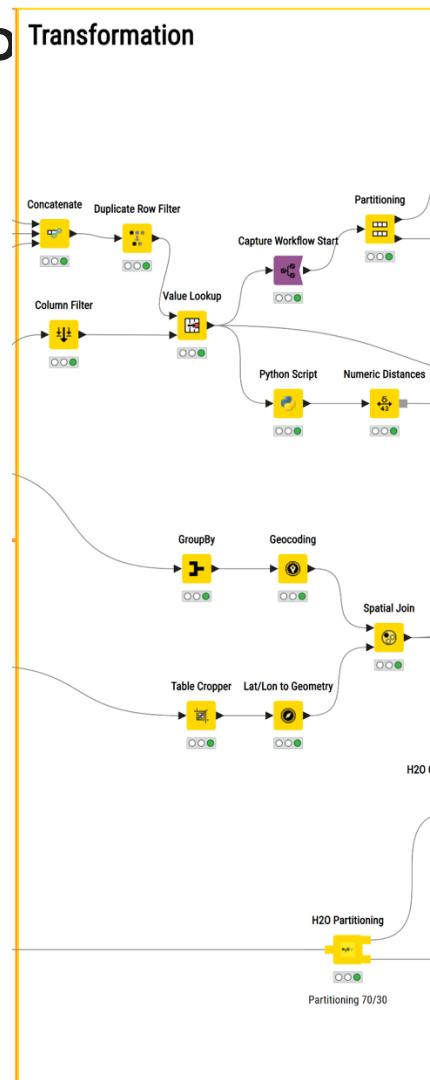
Big Data

Big Data: Loading and pre-processing



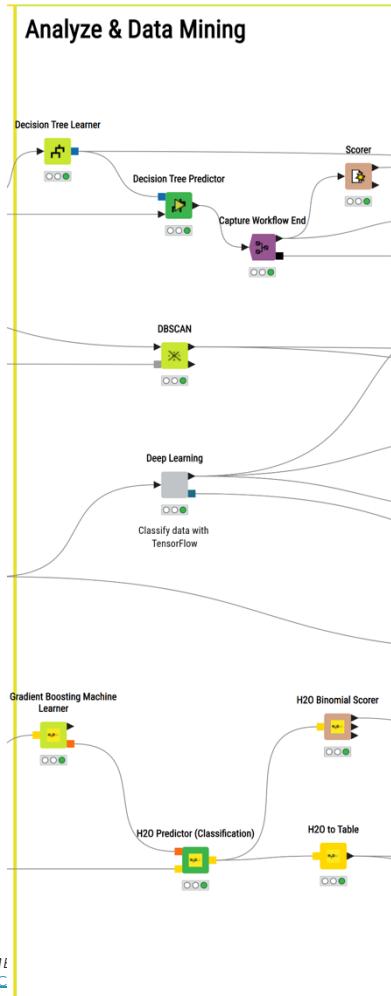
- Spark & Databricks
- HDFS support
- Hive
- Impala
- In-database processing

Transformation



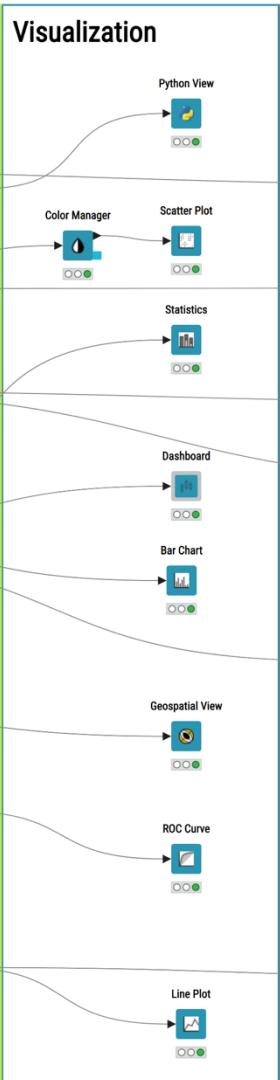
- Data Cleaning
 - Row, column, table based
- Data Merging
 - Join, concatenate, append
- Aggregation
 - Grouping, pivoting, binning
- Feature creation and selection

Analyze & Data Mining



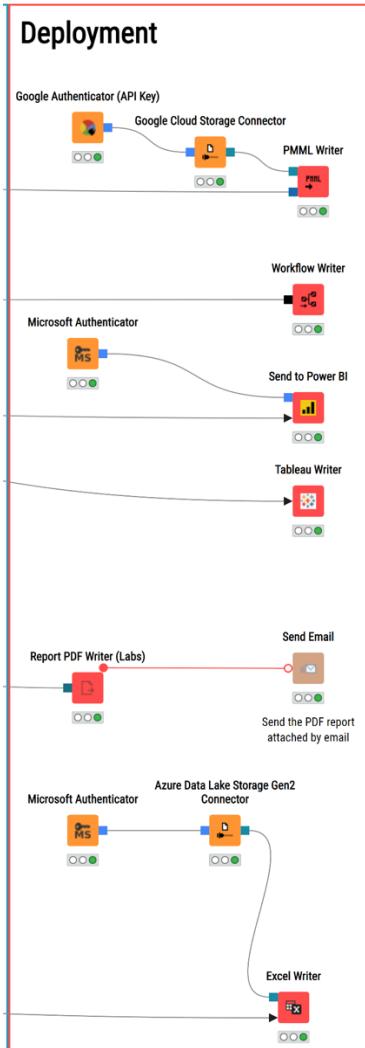
- Regression
 - Linear, regression tree
- Classification
 - Decision tree, ensembles, SVM, MLP, Naïve Bayes, logistic regression
- Clustering
 - k-means, DBSCAN, hierarchical
- Validation
 - Cross-validation, scoring, ROC
- Deep Learning
 - Keras, DL4J
- External
 - R, Python, Weka, H2O, Keras

Visualization



- Interactive dashboards
- Many visualization nodes
 - Scatter Plot, Box Plot, Line Plot
 - Networks, ROC Curve, Decision Tree
 - Adding more with each release!
- Misc
 - Tag cloud, geospatial, molecules
- Script-based visualizations
 - R, Python, JavaScript

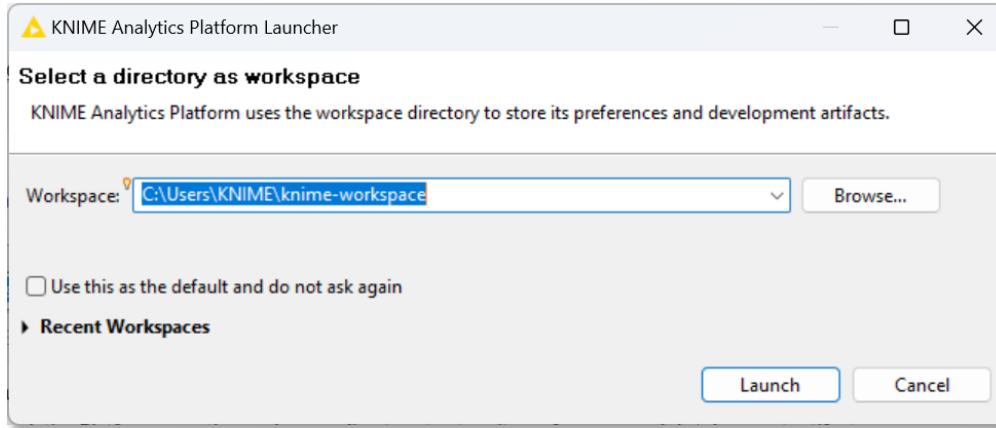
Deployment



- Database
- Files
 - Excel, CSV, txt
 - XML
 - PMML
 - to: local, KNIME Business Hub, Cloud
- Reporting

The KNIME Workspace

- Folder on the local computer in which workflows and data files are stored for the current KNIME session
 - Workspaces are portable, just like KNIME



Tour of the User Interface

The screenshot shows the KNIME Analytics Platform interface. At the top, there's a navigation bar with 'Home', 'Combine Clean and Summarize Spread...', 'Help', 'Preferences', and 'Menu'. Below the navigation bar, there's a 'Recent' section on the left containing 'Local space' and 'KNIME Community Hub' sections, each with small workflow thumbnail previews. A yellow callout bubble points to the 'Local space' section with the text 'Browse local space'. In the center, there's a 'Welcome' section featuring three workflow examples: 'Combine Clean and Summarize Spreadsheet Data' (highlighted with an orange border), 'Countif and Sumif', and 'Non-standard format Spreadsheets'. A yellow callout bubble points to the first example with the text 'Open an example workflow'. Below these examples is a link '→ Show me more examples on the KNIME Community Hub'. To the right of the examples, there's a 'Recently used workflows and components' section listing three items: 'Combine Clean and Summarize Spreadsheet Data' (Local space, just now), 'Non-standard format Spreadsheets' (Local space, 5 minutes ago), and 'Countif and Sumif' (Local space, 5 minutes ago). A yellow callout bubble points to this section with the text 'Create new workflow' and '+ Create new workflow'. At the bottom, a yellow callout bubble points to the 'Recent' section with the text 'Open recent workflows'.

Browse local space

Open an example workflow

Create new workflow

+ Create new workflow

Open recent workflows

Recent

Local space

KNIME Community Hub

Welcome

Dismiss examples X

Combine Clean and Summarize Spreadsheet Data

Countif and Sumif

Non-standard format Spreadsheets

→ Show me more examples on the KNIME Community Hub

Recently used workflows and components

Workflow	Type	Last Used
Combine Clean and Summarize Spreadsheet Data	Local space	just now
Non-standard format Spreadsheets	Local space	5 minutes ago
Countif and Sumif	Local space	5 minutes ago

Tour of the User Interface

The screenshot shows the KNIME User Interface with several UI elements highlighted by yellow callouts:

- Description tab**: A yellow callout pointing to the left sidebar where the "Description" tab is selected. It contains detailed text about the "Excel Reader" node.
- Application tabs**: A yellow callout pointing to the top navigation bar, specifically the "Home" tab.
- Workflow editor**: A yellow callout pointing to the main workspace area where a data processing workflow is displayed. The workflow involves reading Excel files, filtering data, concatenating tables, and calculating volumes.
- Node monitor**: A yellow callout pointing to the bottom right corner where a table titled "1: File Table" displays data from an Excel file.

Workflow Editor Details:

The workflow consists of the following steps:

- Read the first sheet ("Kitchen") of the rooms.xlsx file**: An Excel Reader node.
- Read the "Living Room" sheet of the rooms.xlsx file**: An Excel Reader node.
- Read the "Dinette" sheet of the rooms.xlsx file**: An Excel Reader node.
- Columns Filter**: A node that excludes comments from the "Living Room" sheet.
- Concatenate**: A node that brings items from the "Kitchen" and "Living Room" into one table.
- String to Number**: A node that changes the "Test Volume" column from string to number.
- Value Lookup**: A node that appends volumes of listed items based on their IDs.
- Bar Chart**: A node that creates a bar chart of single item amounts.
- Row Aggregator**: A node that sums up volumes and calculates a grand total.

Node Monitor Data:

#	RowID	Item	Amount
1	Row0	Table	1
2	Row1	Chair	2
3	Row2	Cupboard	1
4	Row3	Side table	1
5	Row4	Fridge	1
6	Row5	Fridge	1

Tour of the User Interface

The screenshot shows the KNIME Analytics Platform interface. On the left, the **Node Repository tab** is selected, displaying various data source and manipulation nodes like Excel Reader, Google Sheets Reader, Row Filter, and Value Lookup. A yellow callout points to the **Search and browse nodes** button at the top of the repository view.

In the center, a yellow callout points to the **Drag and drop a node to add it to the workflow** button, located above the workflow canvas. The workflow canvas itself shows a task for calculating furniture volume, using nodes like Read from file, Column Filter, String to Number, and Bar Chart.

At the bottom, a data preview table titled "1: File Table" shows a list of rows and columns, with a "Flow Variables" section above it.

```
graph LR; subgraph Workflow [Workflow]; Read1[Read the first sheet "Kitchen" of the rooms.xlsx file] --> ColFilter1[Column Filter]; Read2[Read the "Living Room" sheet of the rooms.xlsx file] --> ColFilter2[Column Filter]; Read3[Read the "Dinette" sheet of the rooms.xlsx file] --> String2Num[String to Number]; ColFilter1 --> ColMerge1[Column Merger]; ColFilter2 --> ColMerge1; ColMerge1 --> String2Num; String2Num --> Charge[Change "Test Volume" column from string to number]; Charge --> Concatenate[Concatenate]; ColMerge2[Column Merger] --> Charge; ColMerge2 --> ValueLookup1[Value Lookup]; ValueLookup1 --> Append[Append the volumes of the listed items based on their IDs]; end; Concatenate --> BarChart[Bar Chart]; Append --> BarChart;
```

#	RowID	Item String	Amount
1	Row0	Table	1
2	Row1	Chair	2
3	Row2	Cupboard	1
4	Row3	Side table	1
5	Row4	Fridge	1
6	Row5	Fridge	1

Tour of the User Interface

The screenshot shows the KNIME Analytics Platform interface. On the left, a yellow callout points to the "Space Explorer tab" icon in the sidebar, which is highlighted with a red circle. The main workspace displays a workflow titled "Combine Clean and Summarize Spread". A large yellow callout in the center says "Browse workflows in your local space". Below the workspace, there is a table with 9 rows and 2 columns:

#	RowID	Item	Amount
1	Row0	Table	1
2	Row1	Chair	2
3	Row2	Cupboard	1
4	Row3	Side table	1
5	Row4	Fridge	1
6	Row5	Fridge	1

Tour of the User Interface

The screenshot shows the KNIME Analytics Platform interface. On the left, there's a sidebar with icons for Home, Recent, Execute, Cancel, and Reset. The main area is titled "Combine Clean and Summarize Spread..." and shows a workflow titled "Combine Clean and Summarize Spreadsheet".

Workflow Monitor tab: A yellow callout points to the "Workflow monitor" tab in the sidebar. Another yellow callout points to the "Node errors" section, which displays an error message: "Execute failed: The specified file Rooms1.xlsx does not exist." An orange box highlights the "x" icon next to the error message.

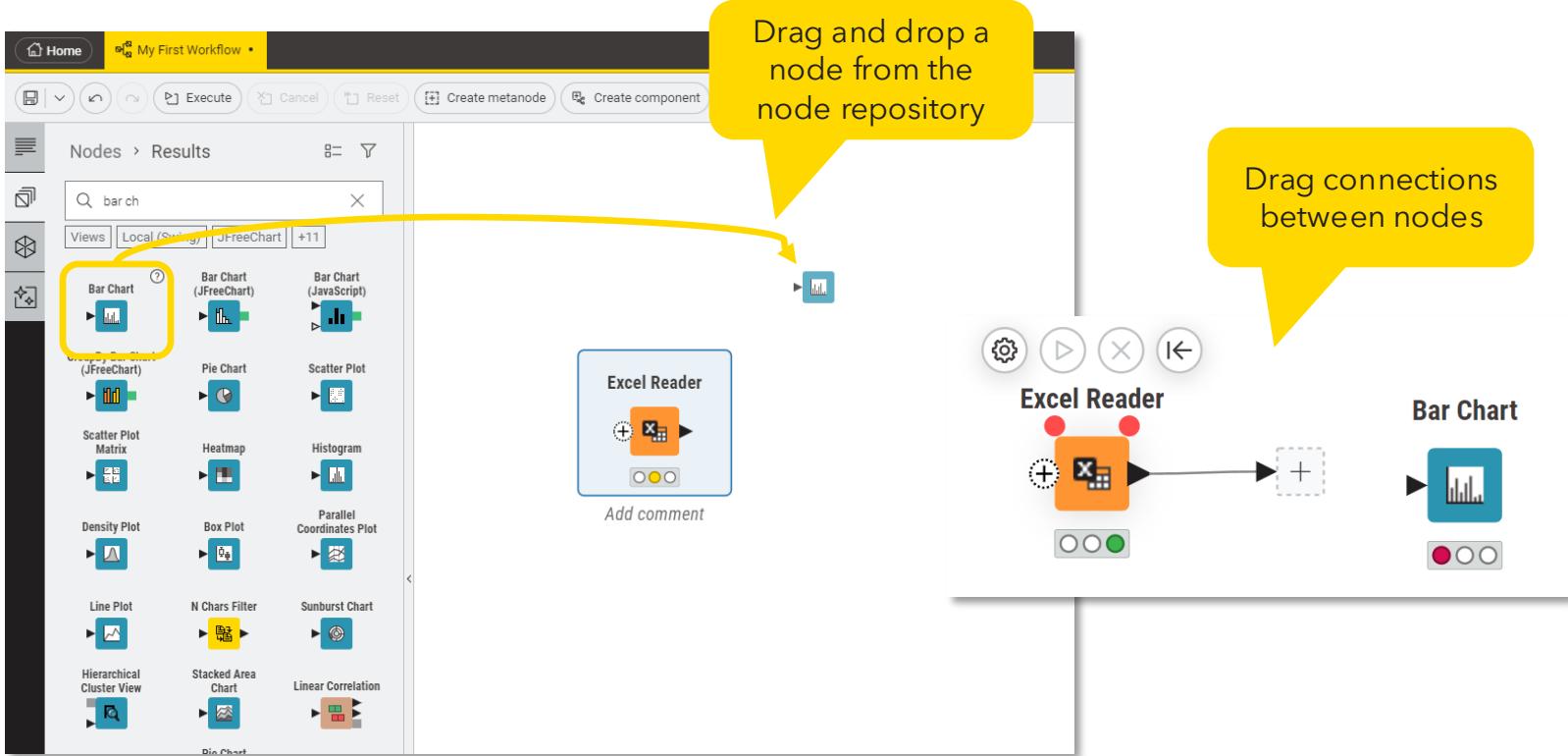
List error and warnings in the workflow: A yellow callout points to the "Node errors" section, which also lists other errors and warnings such as "shifting data that was entered into a wrong column" and "changing the data type of a column".

Click to highlight the failing node in the workflow: A yellow callout points to a speech bubble containing the text "Click to highlight the failing node in the workflow". This bubble is positioned over the workflow canvas where several nodes are connected by arrows.

Workflow Canvas: The central part of the interface shows the workflow. It starts with "Excel Reader" nodes reading data from "Rooms1.xlsx" and "Rooms2.xlsx". These are followed by "Columns Filter", "Exclude the comments from the 'Living Room' sheet", "String to Number", and "Change 'test Volume' column from string to number" nodes. The data then flows through a "Concatenate" node, which is highlighted with a yellow box. After concatenation, it goes through a "Value Lookup" and "Append the volumes of the listed items based on their IDs" nodes. Finally, the data is visualized using a "Bar Chart" and a "Row Aggregator" node.

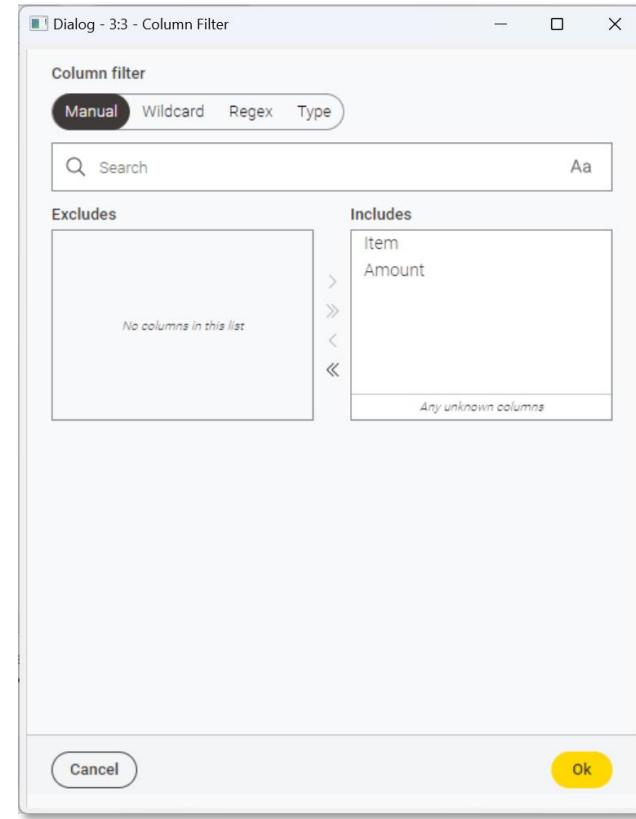
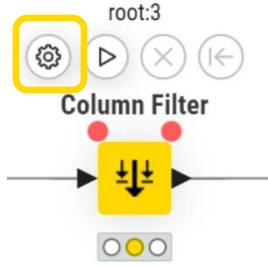
Bottom Panel: The bottom panel has tabs for "1: File Table" and "Flow Variables". It includes a message: "To show the port output, please execute the selected node." and a "Execute" button.

Add and Connect Nodes



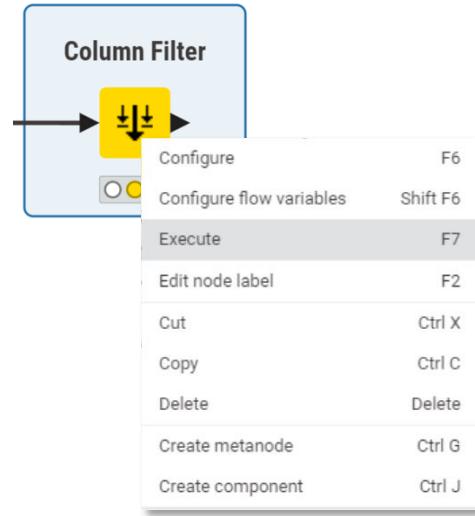
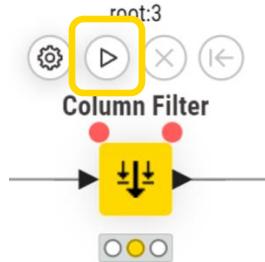
Configure a Node

- Most nodes require configuration
- Options to open the configuration dialog of a node:
 - Double click the node
 - Right click -> Configure
 - Select the node and press F6
 - Click the Configure button the node action bar



Execute a Node

- An executed node has carried out its task
- Options to execute a node:
 - Right click -> Execute
 - Select the node and press F7
 - Click the Execute button the node action bar



Node Status

- A node can have 4 statuses:

Excel Reader



Not Configured:

The node is waiting for configuration or incoming data.

Excel Reader



Configured:

The node has been configured correctly and can be executed.

Excel Reader



Executed:

The node has been successfully executed.

Results may be viewed and used in downstream nodes.

Excel Reader

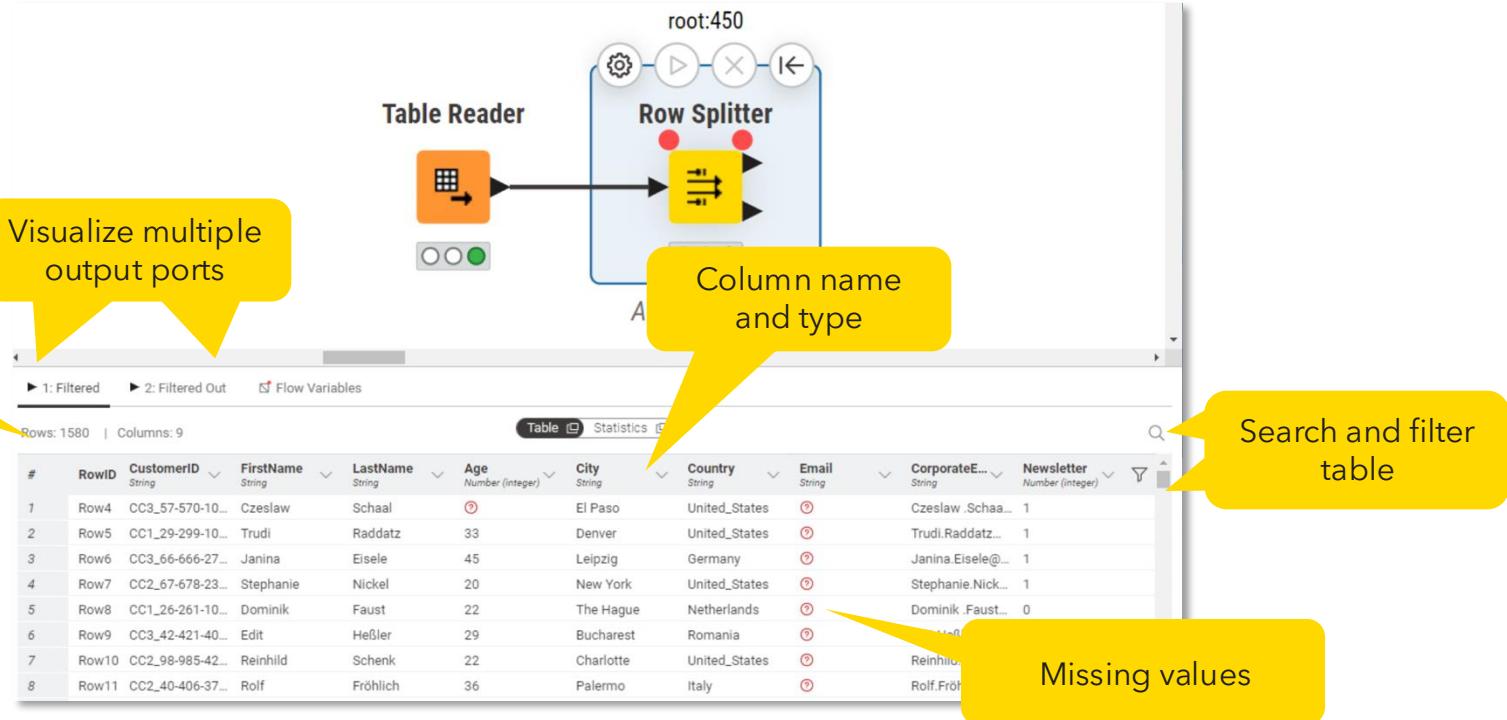


Error:

The node has encountered an error during execution.

Node Monitor

- Select an executed node to inspect its output(s)



Statistics Tab

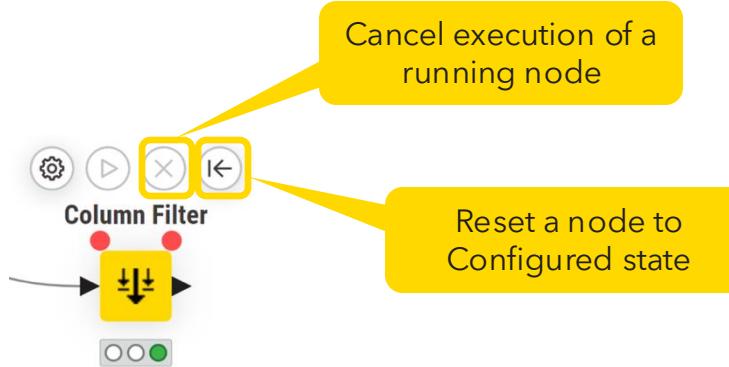
- Get statistics of the selected node output port
 - Missing values
 - Unique values
 - Min and max
 - Quantiles
 - ...

The screenshot shows the KNIME interface with the following details:

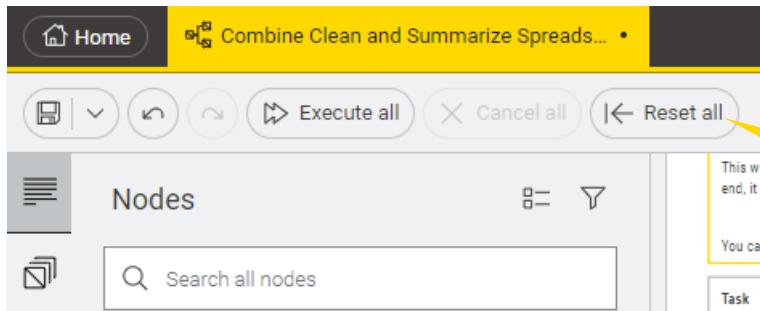
- Node:** Excel Reader (highlighted with a blue box).
- Add comment:** A button below the node icon.
- Flow:** Shows "1: File Table" and "Flow Variables".
- Data View:** Shows a table with 3 rows and 14 columns. The columns are labeled: Name, Type, # Missing v..., # Unique val..., Minimum, Maximum, 25% Quantile, 50% Quantil..., 75% Quantile, Mean.
- Statistics Tab:** A yellow speech bubble points to the "Statistics" tab in the bottom navigation bar of the data view.
- Data Preview:** The table data is as follows:

Name	Type	# Missing v...	# Unique val...	Minimum	Maximum	25% Quantile	50% Quantil...	75% Quantile	Mean
Continent	String	0	7	?	?	?	?	?	?
Population (2...	Number (long)	0	7	0	4,753,079,726	46,004,866	604,182,517	1,460,481,772	1,149,128
Area	Number (inte...	0	7	3,103,313	29,648,481	8,486,460	17,461,112	22,134,710	16,554,86

Cancel and Reset a Node



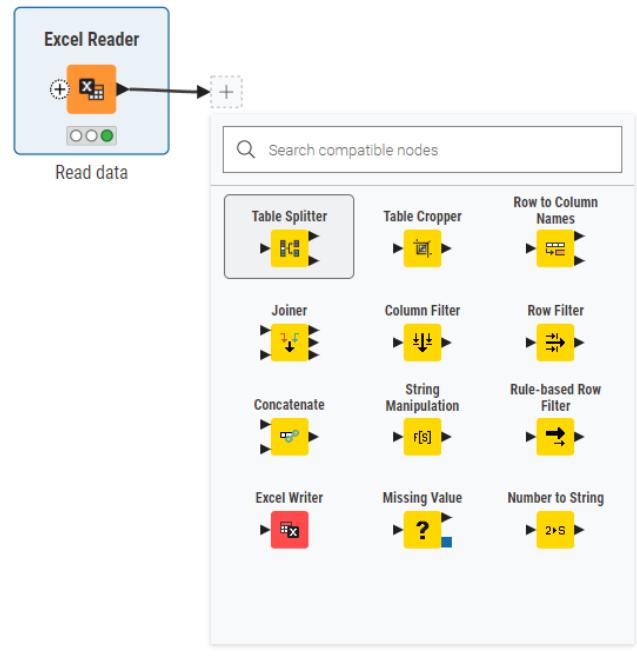
- Node control is also possible from the Workflow Toolbar



If no node is selected,
you can control all
nodes in the workflow

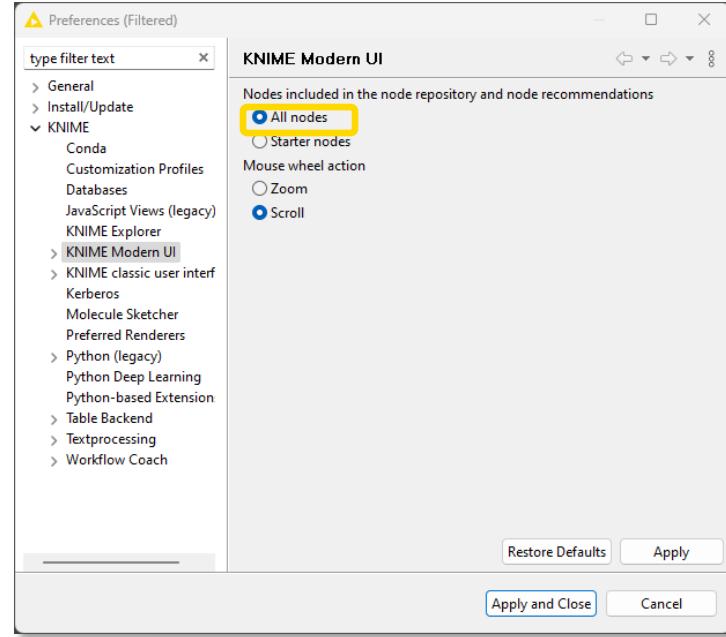
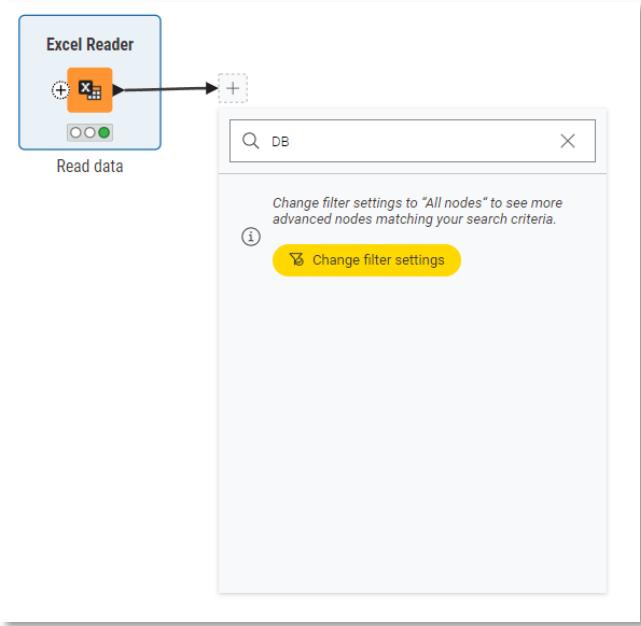
Quick Node Adding Panel

- Drag and drop the output of a node to an empty space
- See recommended nodes to append
- Search for all compatible nodes



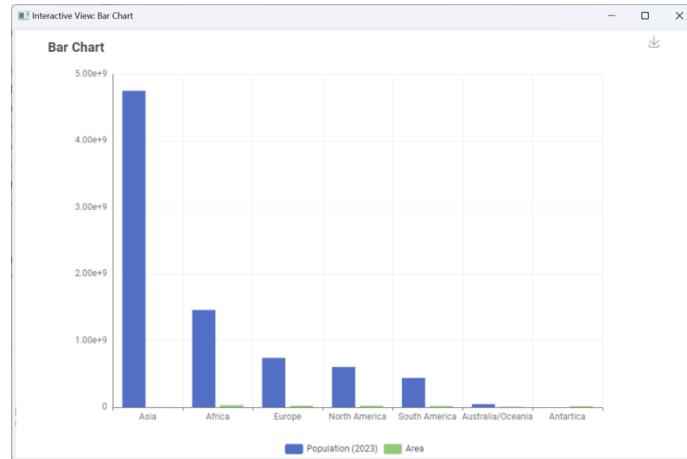
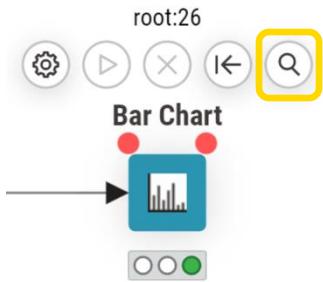
Starter Perspective

- Click “Change filter settings” and select “All nodes” to see the full list
- By default, only a curated set of essential nodes is displayed



Node Views

- Open the node view in a new window
- Only available for “blue” nodes



KNIME File Extensions

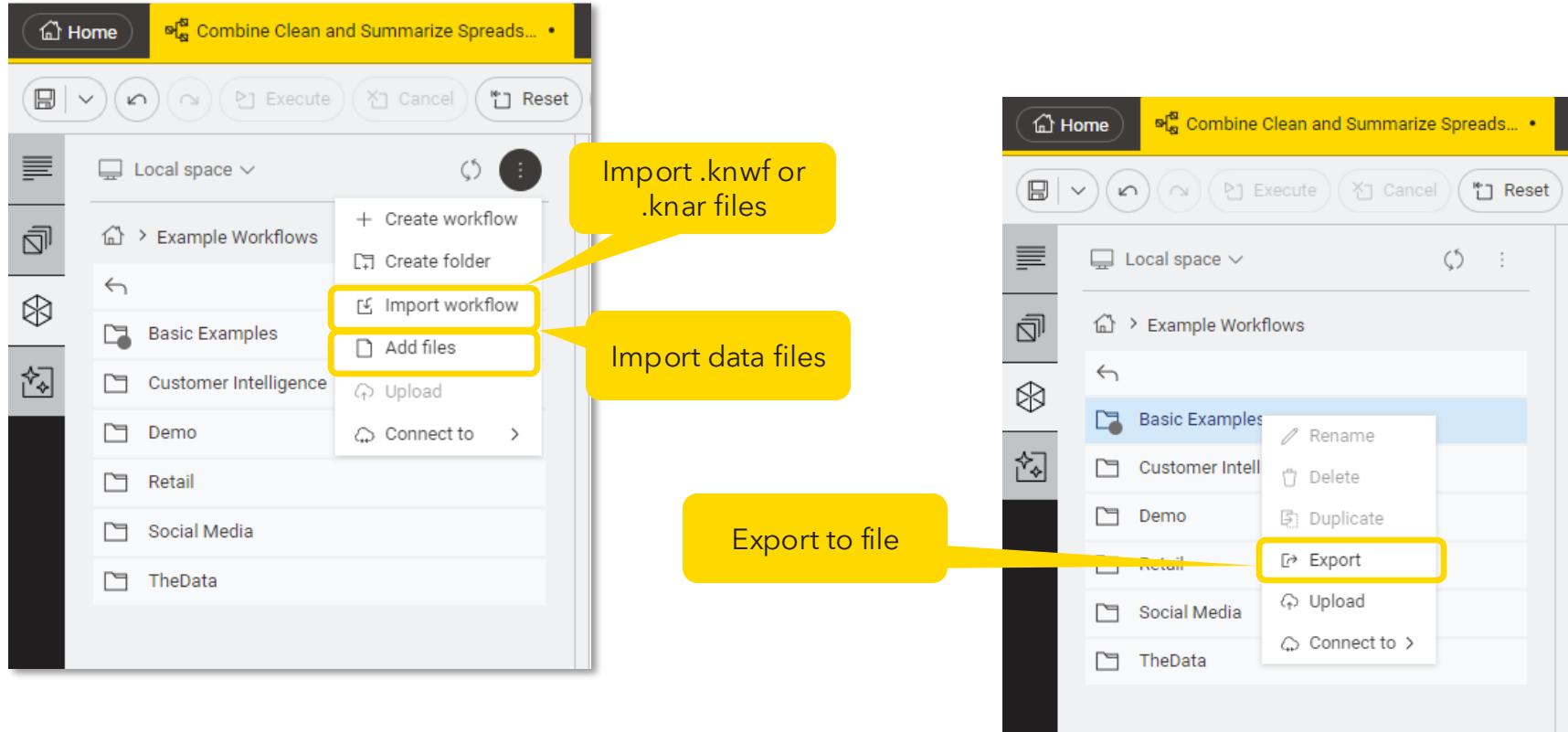
- Dedicated file extensions for workflows and workflow groups associated with KNIME Analytics Platform
- ***.knwf** for KNIME Workflow Files



- ***.knar** for KNIME Archive Files (folder)



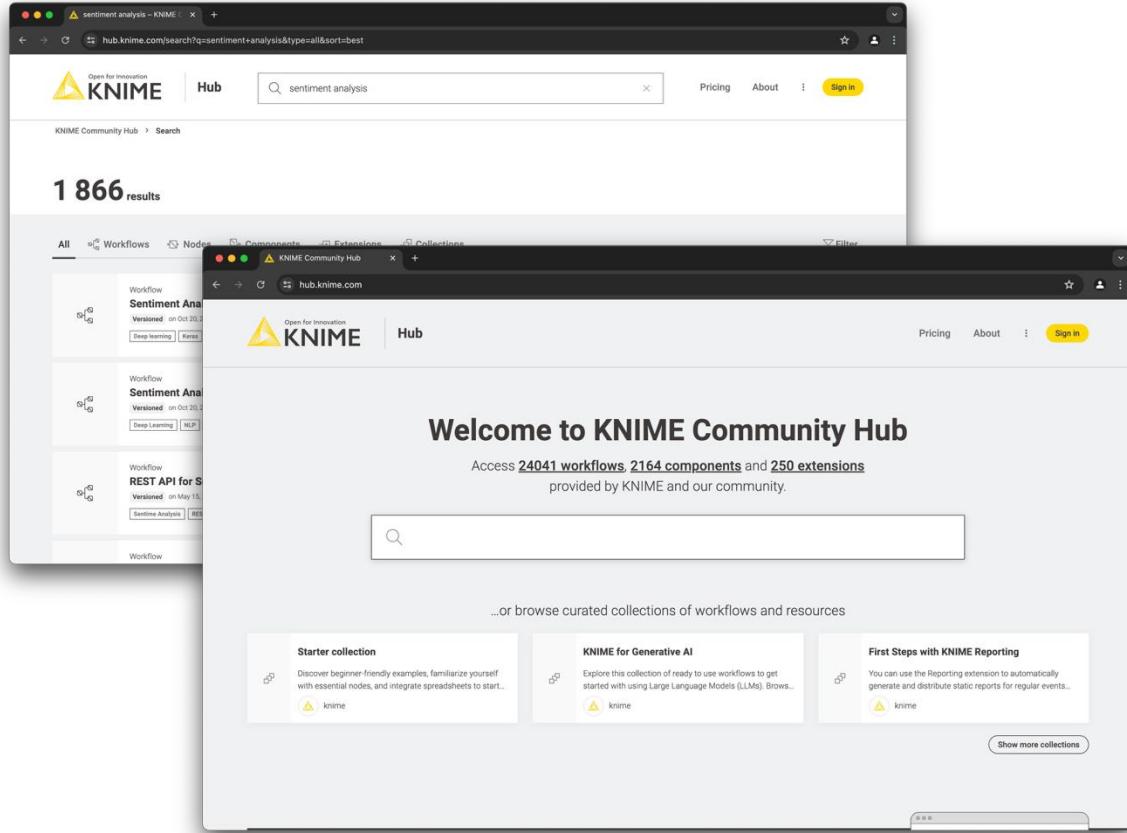
Importing and Exporting Workflows



KNIME Community Hub

- Place to search and share
 - Workflows
 - Nodes
 - Components
 - Extensions
- Collaborate in teams
- Execute and deploy workflows

hub.knime.com



Downloading from KNIME Community Hub

- Sign-in is required to download a folder
- Single workflows can be downloaded without login

The screenshot shows the KNIME Community Hub interface. At the top, there's a navigation bar with the KNIME logo, a search bar, and user profile icons. Below the header, the URL indicates the user is in the 'Education' space under 'Learnathons' for 'Teens Crunch Data The Bootcamp'. A message bubble on the right says 'You must be signed in'. The main content area displays a list of workflow files:

- Day1_READ
- Day2_DISPLAY
- Day3_CONTROL
- Day4_LOOP
- Example Guess The Song Game
- data

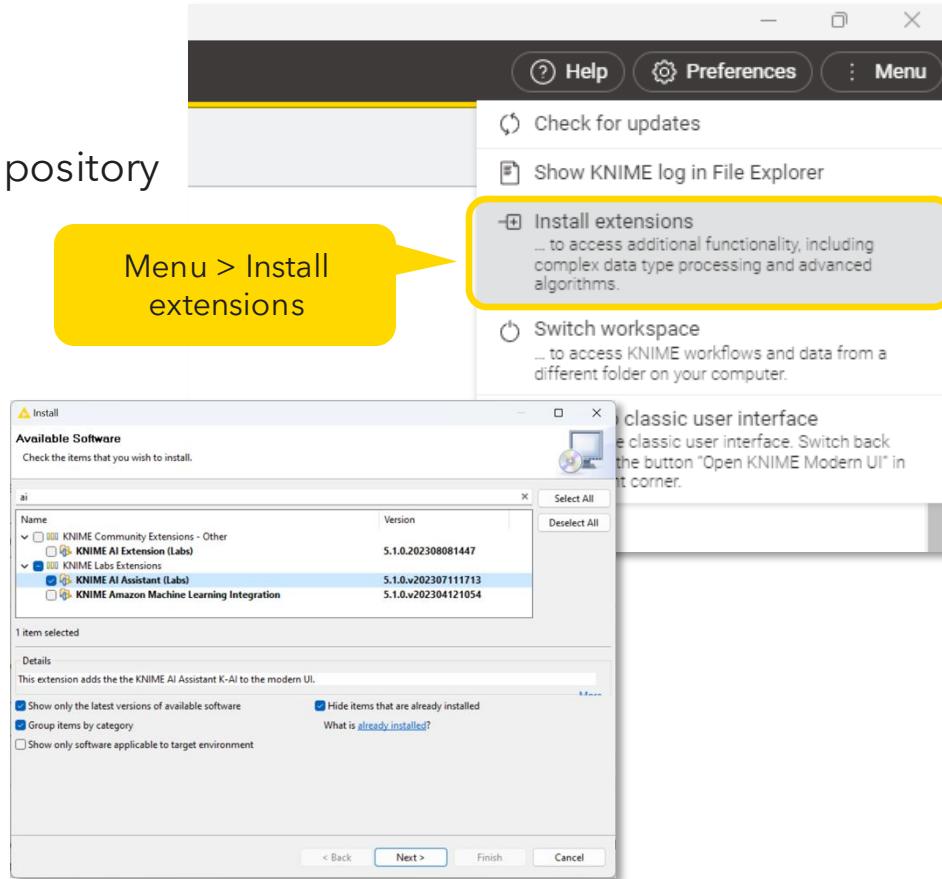
Each file has a download icon (a downward arrow) to its right. A yellow callout points to the download icon for the 'data' file, with the text 'Download full .knar'.

You must be signed in

Download full .knar

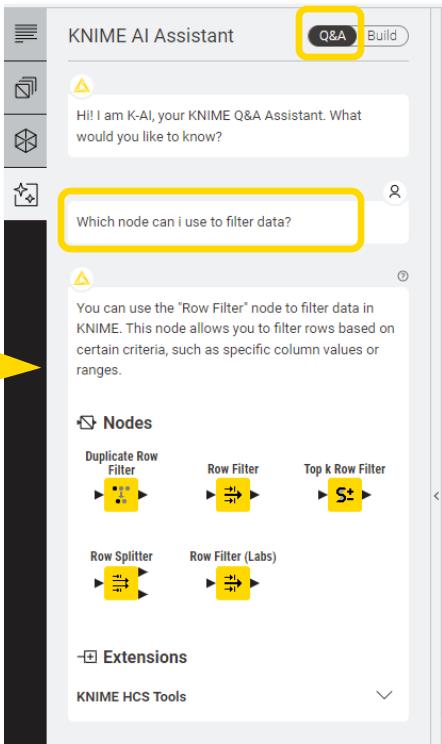
KNIME Extensions

- Provide additional functionalities
- Add more nodes to the Node Repository
 - Text Processing
 - Image Processing
 - Deep Learning
 - Geospatial Analytics
 - Big Data Connectors
 - and many more

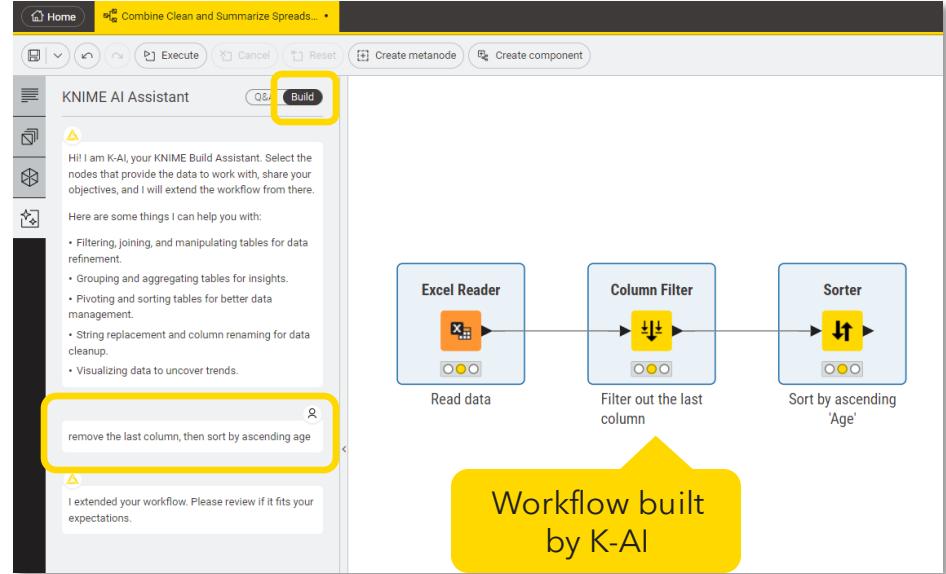


KNIME AI Assistant

- AI-powered chat bot

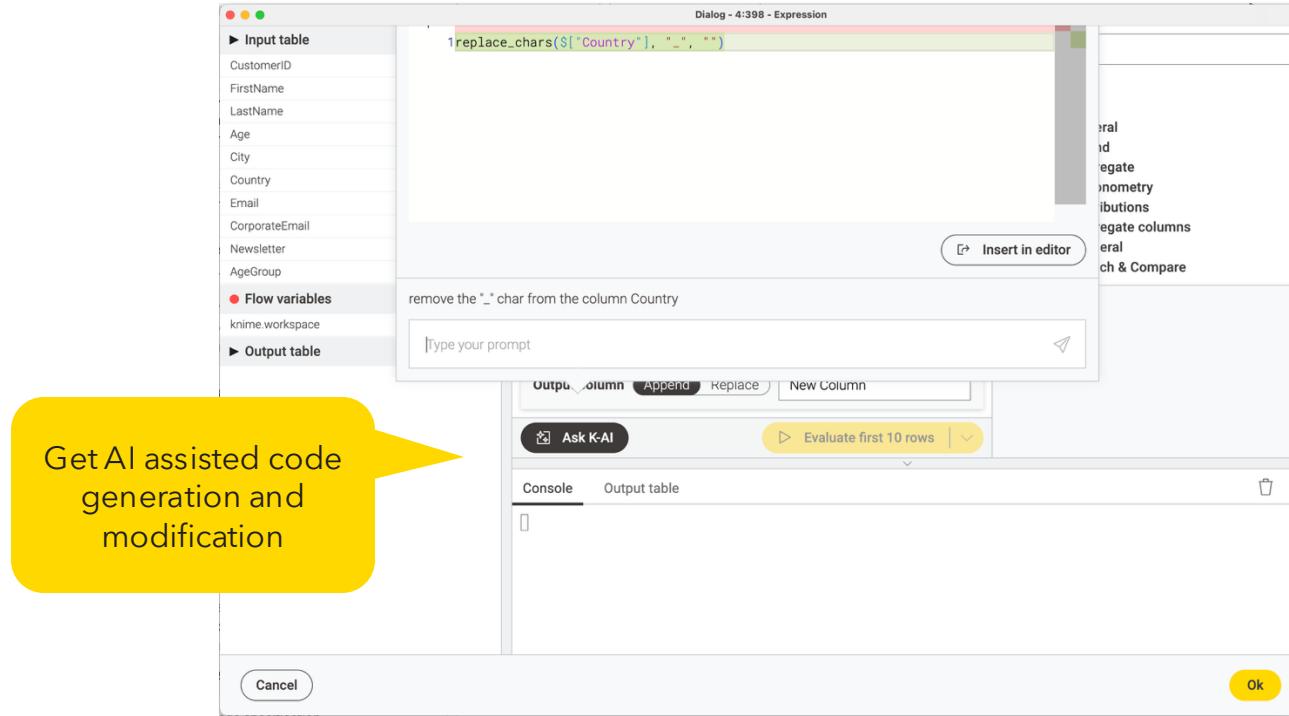


- AI-powered workflow builder



KNIME AI Assistant

- Copilot in scripting nodes



Build Your First Workflow

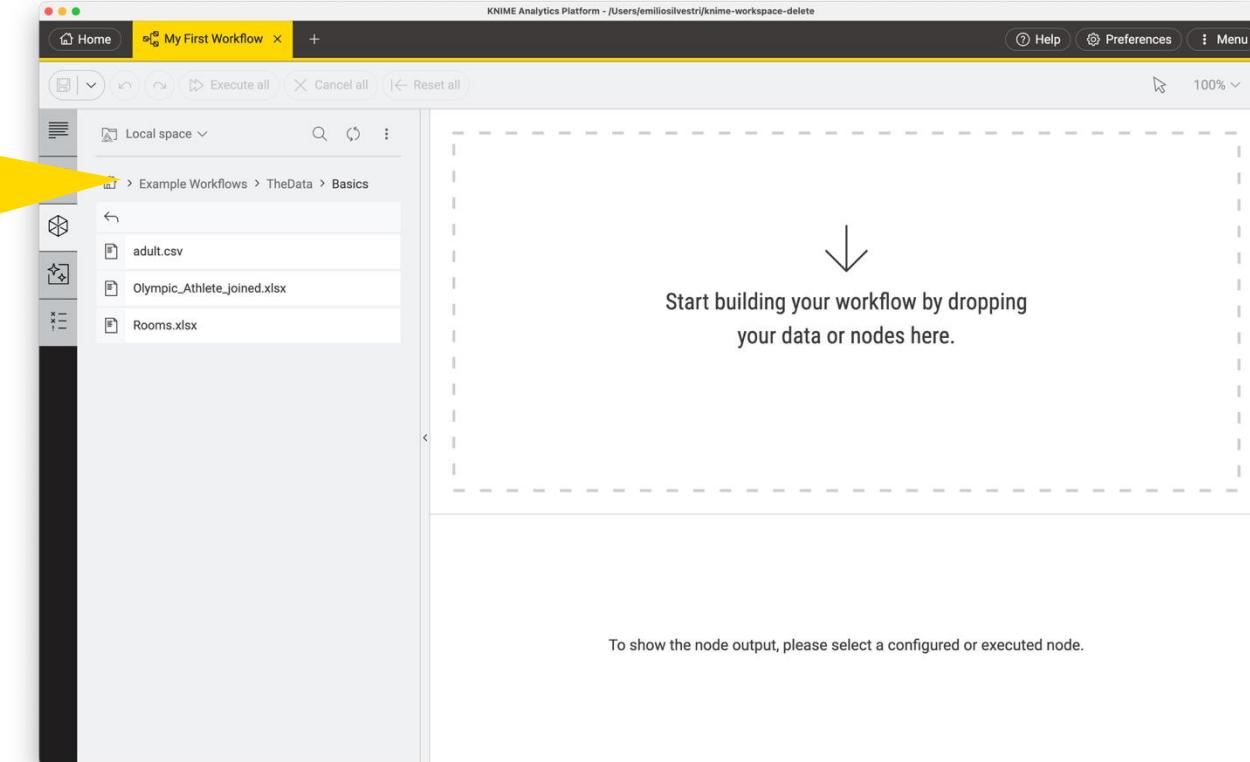
Build Your First Workflow

- From the Home page, create a new workflow and give it a name

The screenshot shows the KNIME Analytics Platform interface. At the top, the title bar reads "KNIME Analytics Platform - /Users/emiliosilvestri/knime-workspace-delete". The main area is titled "Welcome" and features three examples: "Combine Clean and Summarize Spreadsheet Data", "Countif and Sumif", and "Non-standard format Spreadsheets". Below these examples is a link "→ Show me more examples on the KNIME Community Hub". On the left, there's a sidebar with sections for "Recent", "Local space", and "KNIME Community Hub", each showing thumbnail previews of workflows. A call-to-action "Create your KNIME account" is present with a "Create Account" button. On the right, a section titled "Recently used workflows and components" lists three items: "Combine Clean and Summarize Spreadsheet Data" (Local space, just now), "Non-standard format Spreadsheets" (Local space, 5 minutes ago), and "Countif and Sumif" (Local space, 5 minutes ago). A prominent yellow rectangular box highlights the "Create new workflow" button, which is located at the bottom right of the recently used section.

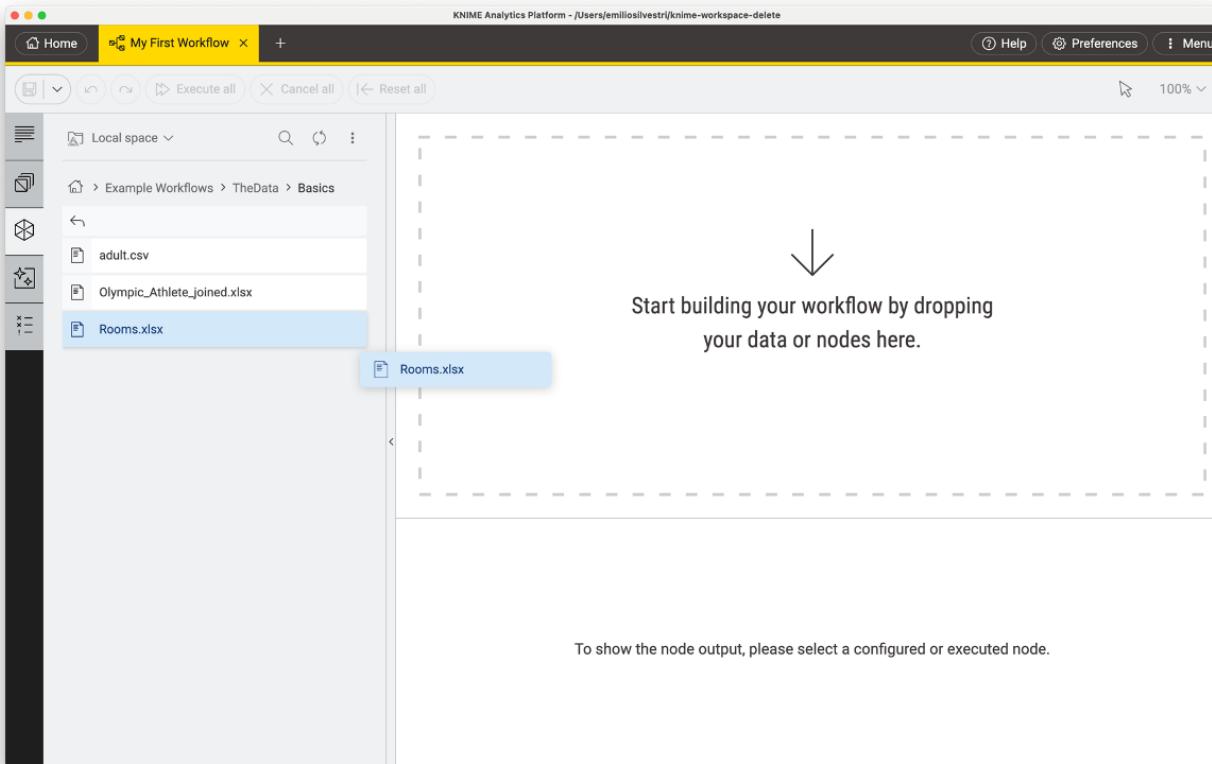
Build Your First Workflow

- From the side panel navigation, open the Space Explorer tab



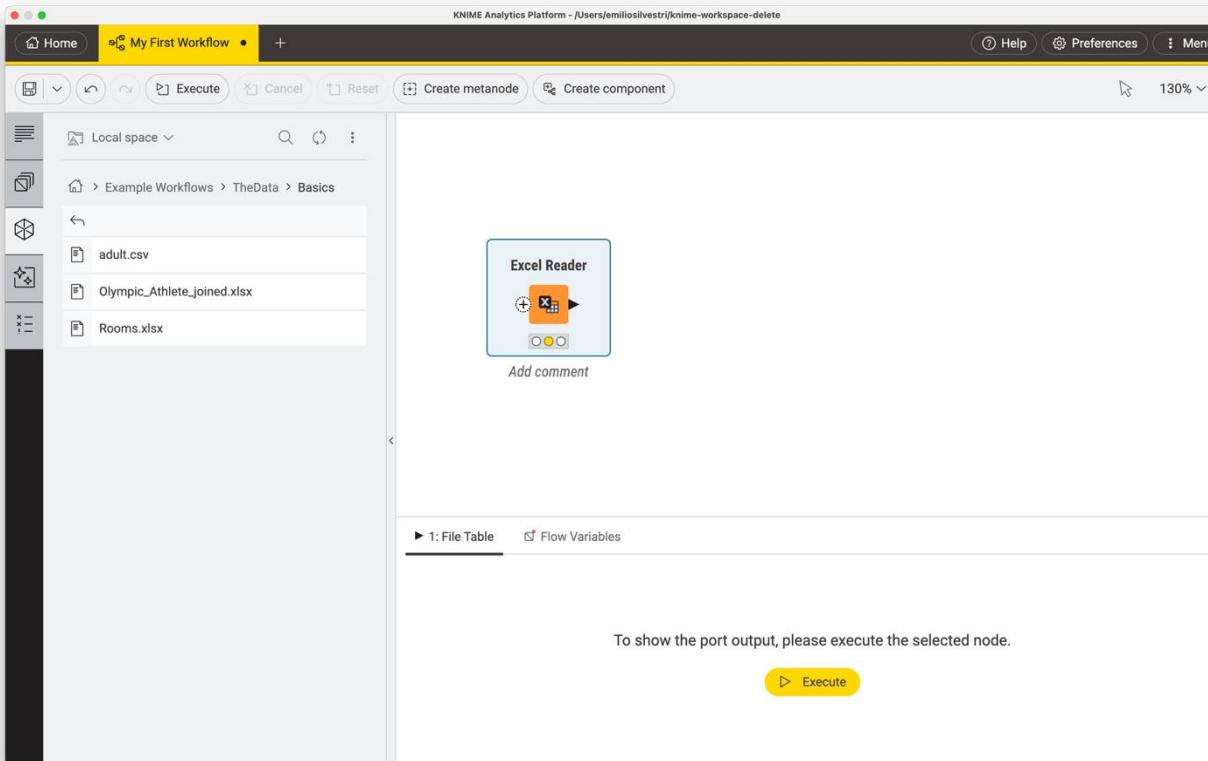
Build Your First Workflow

- Drag and drop the file Rooms.xlsx into the empty Workflow Editor



Build Your First Workflow

- An Excel Reader node appears in the Workflow Editor



Build Your First Workflow

- Select the node and execute it

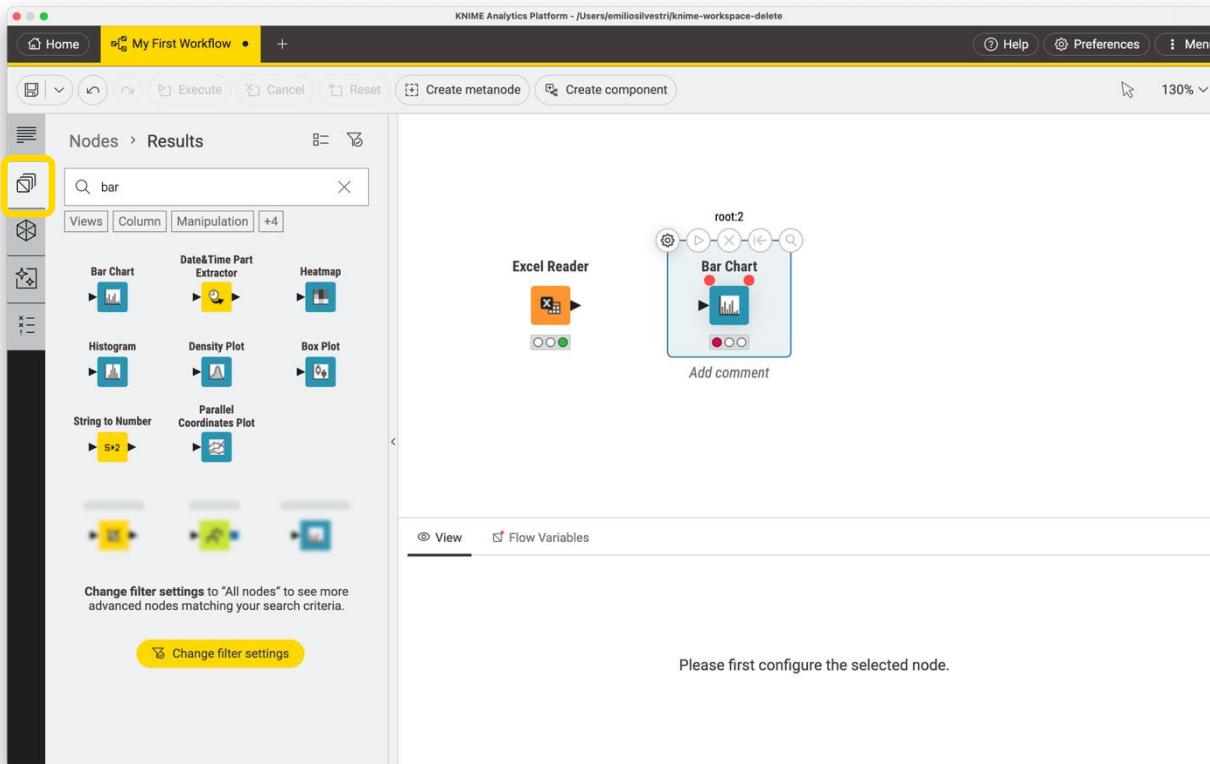
The screenshot shows the KNIME Analytics Platform interface. On the left, the project tree displays files like 'adult.csv', 'Olympic_Athlete_joined.xlsx', and 'Rooms.xlsx'. The main workspace contains a single node, 'root:1', which is an 'Excel Reader' node. A yellow callout bubble points to the play button on the node's toolbar with the text: 'Press the play button to execute the node'. Below the workspace is a 'File Table' node monitor. The table has columns: RowID, Item, and Amount. It contains 9 rows of data:

RowID	Item	Amount	
1	Row0	Table	1
2	Row1	Chair	2
3	Row2	Cupboard	1
4	Row3	Side table	1
5	Row4	Fridge	1
6	Row5	Freezer	1
7	Row6		
8	Row7		
9	Row8		

A yellow callout bubble points to the table with the text: 'Inspect the output of the node in the node monitor'.

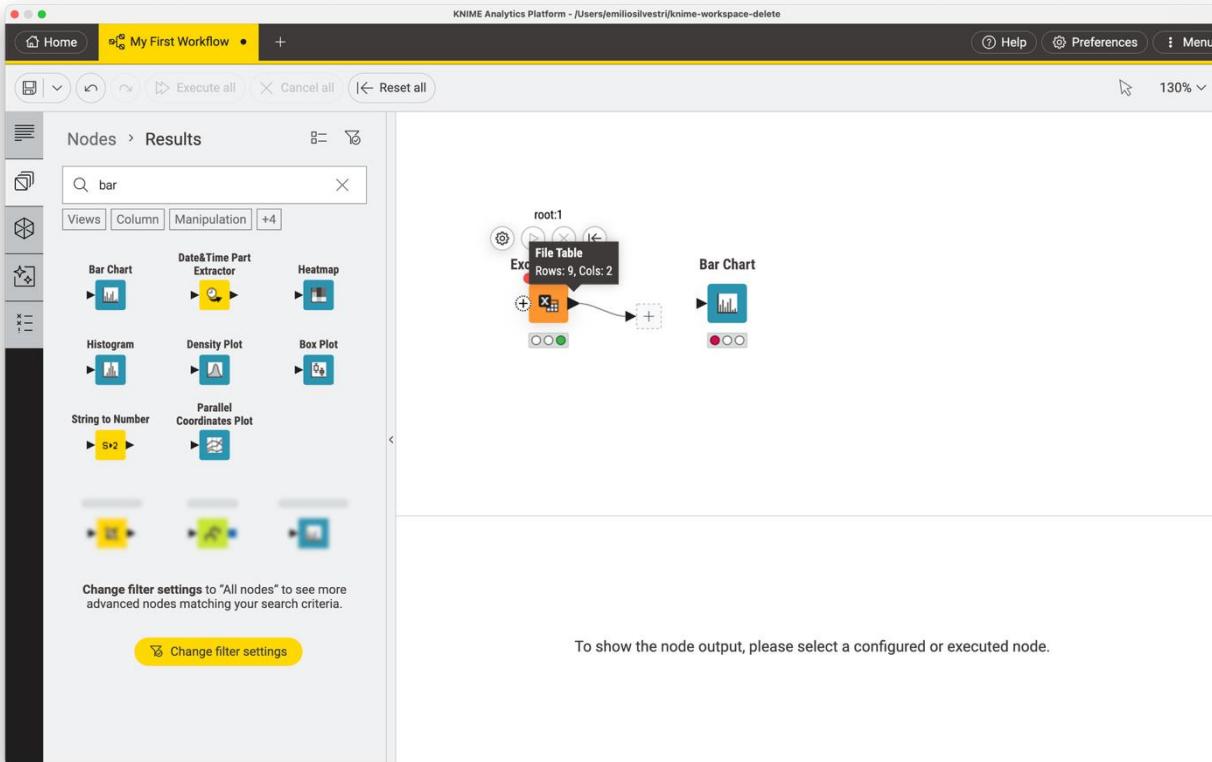
Build Your First Workflow

- From the Node Repository tab, drag and drop the Bar Chart node



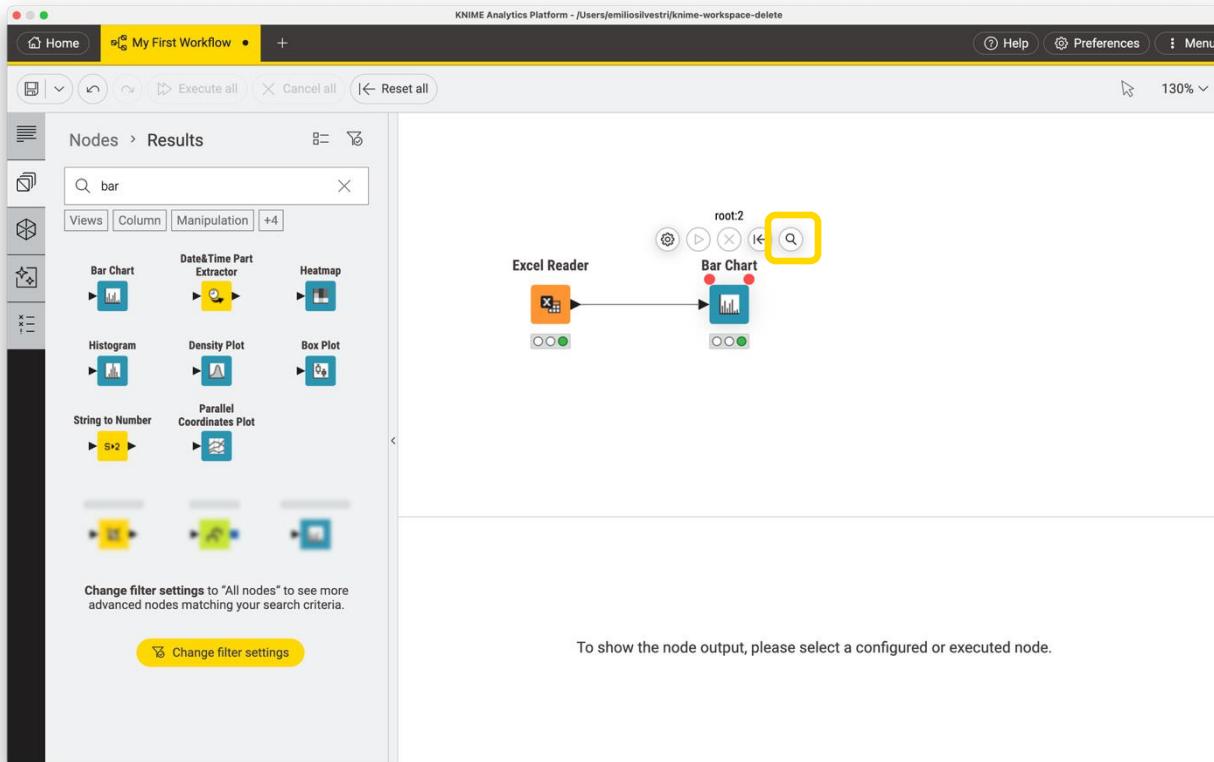
Build Your First Workflow

- Drag the connection between the two nodes



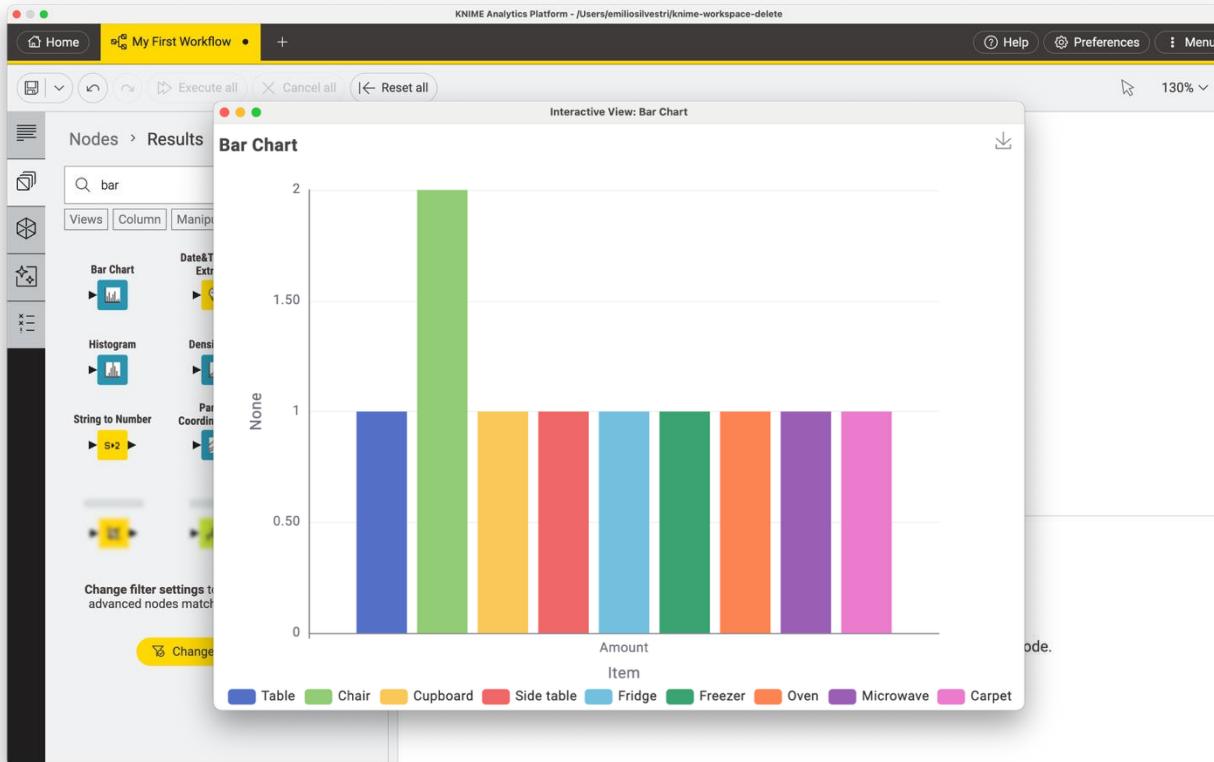
Build Your First Workflow

- Hover over the Bar Chart and click the magnifier icon

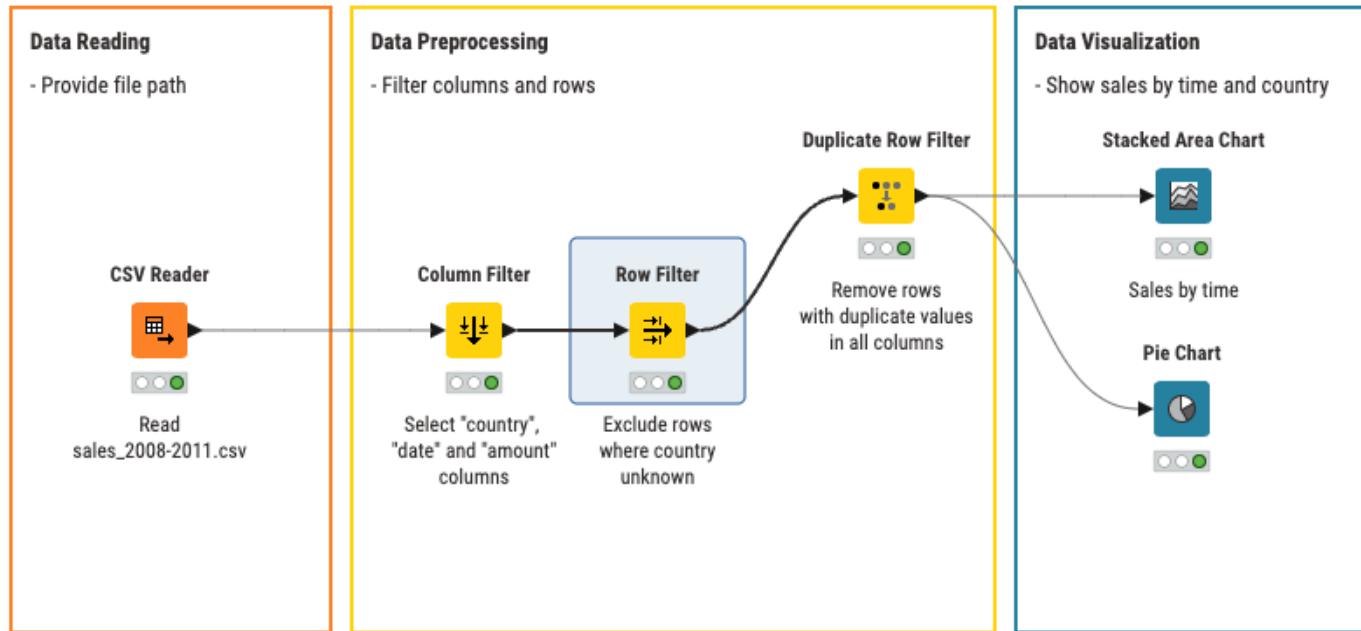


Build Your First Workflow

- Have a look at your first data visualization with KNIME Analytics Platform!



Atividade 1



Atividade 2: introdução

- A mineração de dados é a etapa do processo de descoberta de conhecimento responsável pela seleção dos métodos a serem utilizados para localizar padrões nos dados.
- A identificação destes padrões é uma das possibilidades de análise de dados existentes.
- Dentre as tarefas de mineração de dados, são citadas a clusterização, a associação, a classificação e a regressão.
- Considerando o dataset Iris, um popular conjunto de dados que consiste em 150 amostras das três espécies de flores Iris (setosa, virginica e versicolor), sendo 50 amostras de cada espécie.
- Os dados sobre comprimento e la



largura da sépala e

Atividade 2

- Utilizando a plataforma KNIME Analytics, elaborar um fluxo de trabalho (workflow) para realizar a tarefa de mineração de dados regressão no dataset Iris.
- O objetivo é prever qual o valor da largura da pétala da flor a partir das demais variáveis (Espécie, comprimento e largura da sépala e comprimento da pétala).
- Após a criação e execução do workflow, responda às questões sobre os resultados encontrados.

Atividade 2

Mineração de Dados - Regressão (Prever valor da largura da pétala)



Atividade 2

- Neste fluxo de trabalho, vamos fazer a mineração de dados, utilizando a técnica de regressão.
- Para isso, vamos usar a técnica de regressão linear. O nó de particionamento deve ser configurado da mesma forma que na atividade 1.
- Em seguida, devemos usar os nós “Linear Regression Learner” e “Regression Predictor”, para respectivamente criar o modelo de predição na amostra de treinamento e executar o modelo na amostra de teste.
- O nó “Numeric Scorer” executa uma avaliação da acurácia do modelo de predição criado. Por fim, construímos um gráfico de barras para visualmente avaliar a acurácia do modelo. Para executar esta atividade, você pode utilizar o arquivo “Analise IRIS - Regressao Linear.knwf”, realizando as devidas alterações para ficar semelhante à figura do workflow apresentada acima.

Atividade 2

