



Introduction to KNIME

Special Topics in Data Analytics (CSIS 4260)

NOUSHIN CHABOKI BEHBAHANI - 300346872
SUDARSAN HARIDAS - 300353099

Introduction

In today's fast-paced, data-driven world, organizations face the challenge of making sense of large and complex datasets. Imagine a tool that combines the power of simplicity with advanced analytics. KNIME (Konstanz Information Miner) is a cutting-edge platform that provides an intuitive environment to explore, manipulate, and transform data. Whether you are a data scientist, a business analyst, or a data science expert, KNIME offers a user-friendly interface that helps you to use the power of data science without needing to know or write complex codes.

With KNIME, you can easily connect different data sources such as spreadsheets or databases. The visual workflow editor allows you to build analytical workflows like a skilled architecture that is designed by data transformations and analyses by drag-and-drop easily. You can clean, filter, visualize, model, or deploy your data and KNIME offers a wide range of built-in nodes and modules.

But KNIME is not just about the tools; it's about the community. You can join a supportive network of data scientists where they share their knowledge, and answer the other's questions. The KNIME includes a lot of ready-to-use workflows, tutorials, and resources that guide you to explore your data.

Availability

KNIME is available for Windows, Mac OS and Linux. KNIME is also a free-to-download open-source tool.

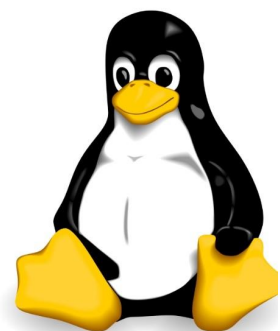
Download link: <https://www.knime.com/downloads>



Microsoft



MacTM OS



Interface Overview

KNIME's User-Friendly Interface

The KNIME interface has a user-friendly design, making it accessible to both beginners and experienced data analysts. It consists of several components that work together to create, manage, and execute data workflows efficiently.

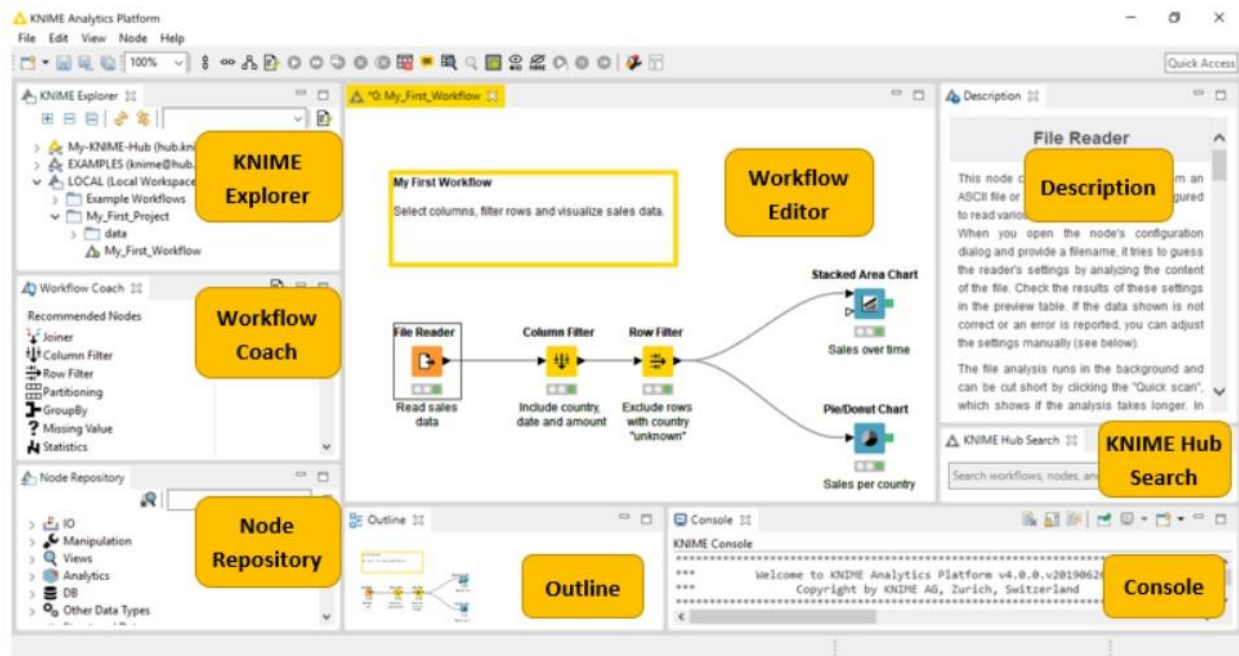


Figure 1 – KNIME workspace.

1. **Workflow Editor:** The central canvas where you build and design your data workflows. You can drag and drop nodes onto this canvas to combine nodes to model data flow.
2. **Node Repository:** Located on the left-hand side, this panel contains a vast library of nodes that perform tasks on data in different parts including IO, manipulation, view, and other types.
3. **Outline View:** This view on the right provides an overview of your workflow's structure. It's useful for navigating large workflows and understanding the flow of data.
4. **Console View:** The console view displays log messages and error outputs, providing feedback on your workflow's execution.
5. **Node Description and Configuration Panel:** When you select a node on the workflow canvas, this panel on the lower left displays information about the selected node and allows you to configure its parameters.

Nodes and workflows

In the workflow, you can represent each task by nodes as shown in Figure 2. Nodes can perform different tasks, including reading or writing files, data transformation, manipulating and model training, visualizations, and so on.

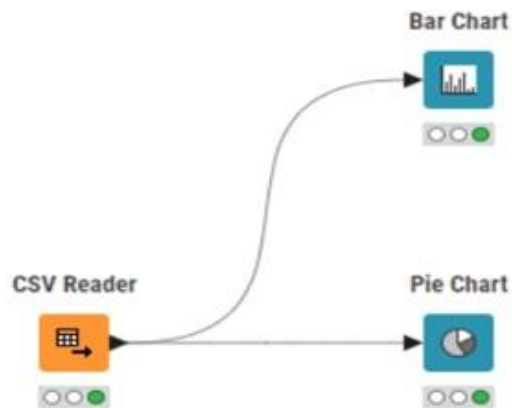


Figure 2- Sample Nodes

The node status is shown by a traffic light as shown in Figure 3.



Figure 3 – Node Status

Each node has specific settings to do a proper configuration. After configuring it, you need to execute that node.

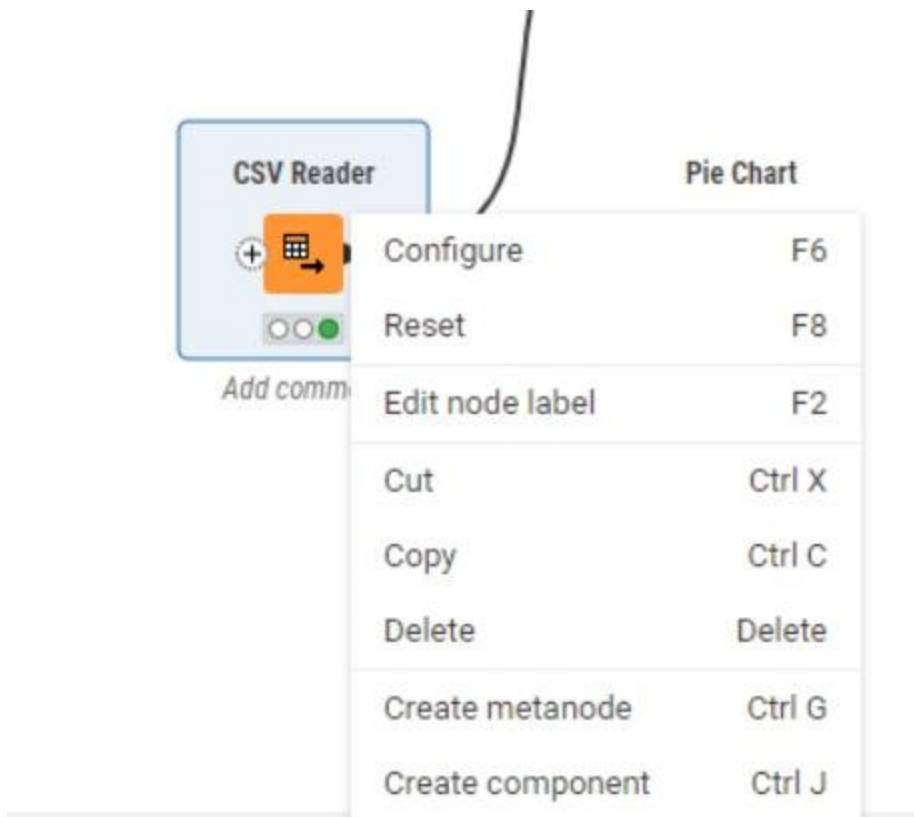


Figure 4 - Node configuration and execution

Comparison with similar workflow tools

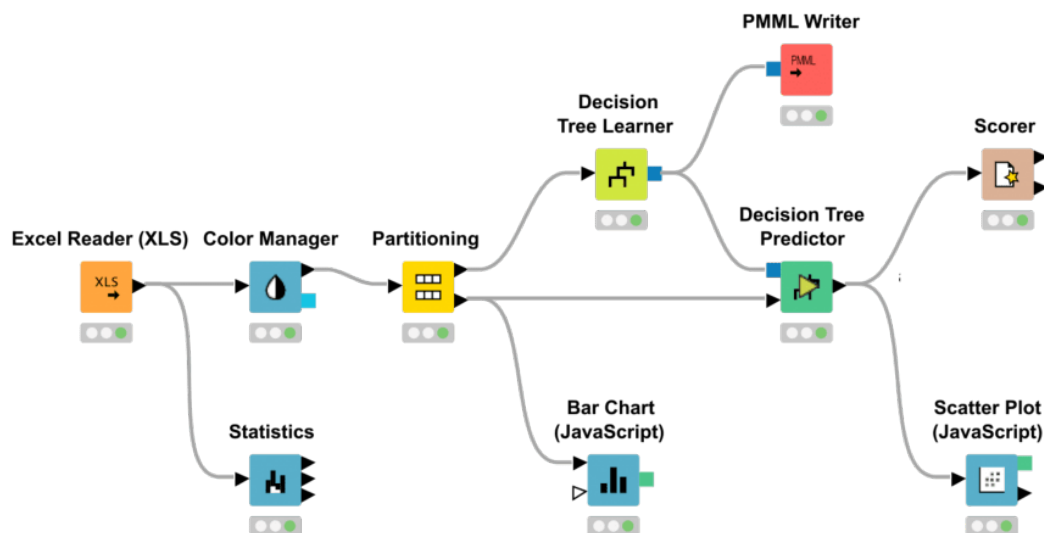
Tools such as Alterxy Designer, Oracle Cloud Data Science Platform, Matlab, Rapidminer Studio, etc., offer similar capabilities. But compared to the others, KNIME is an open-source and free tool and also provides better support and integration across platforms. KNIME is also a no-code tool, so it is easy and faster to integrate KNIME into an organization's needs. KNIME is also known to be very graphics intensive and slow to run.



KNIME challenges and limitations

While KNIME is a widely used data analytics and machine learning platform, like any software, it has its challenges and limitations.

- KNIME can be resource-consumer, especially when dealing with large datasets or complex workflows. It may require considerable memory and processing power, which can be a limitation for users with limited hardware resources.
- For certain complex data processing tasks or machine learning workflows, KNIME's performance is challenging, especially in execution time.
- Although KNIME can connect to big data platforms, handling large volumes of data can be challenging.
- KNIME's capabilities in natural language processing (NLP) and text analytics are not good enough in compared to tools like RapidMiner and users requiring additional plugins or tools.
- KNIME is primarily designed for batch data processing.

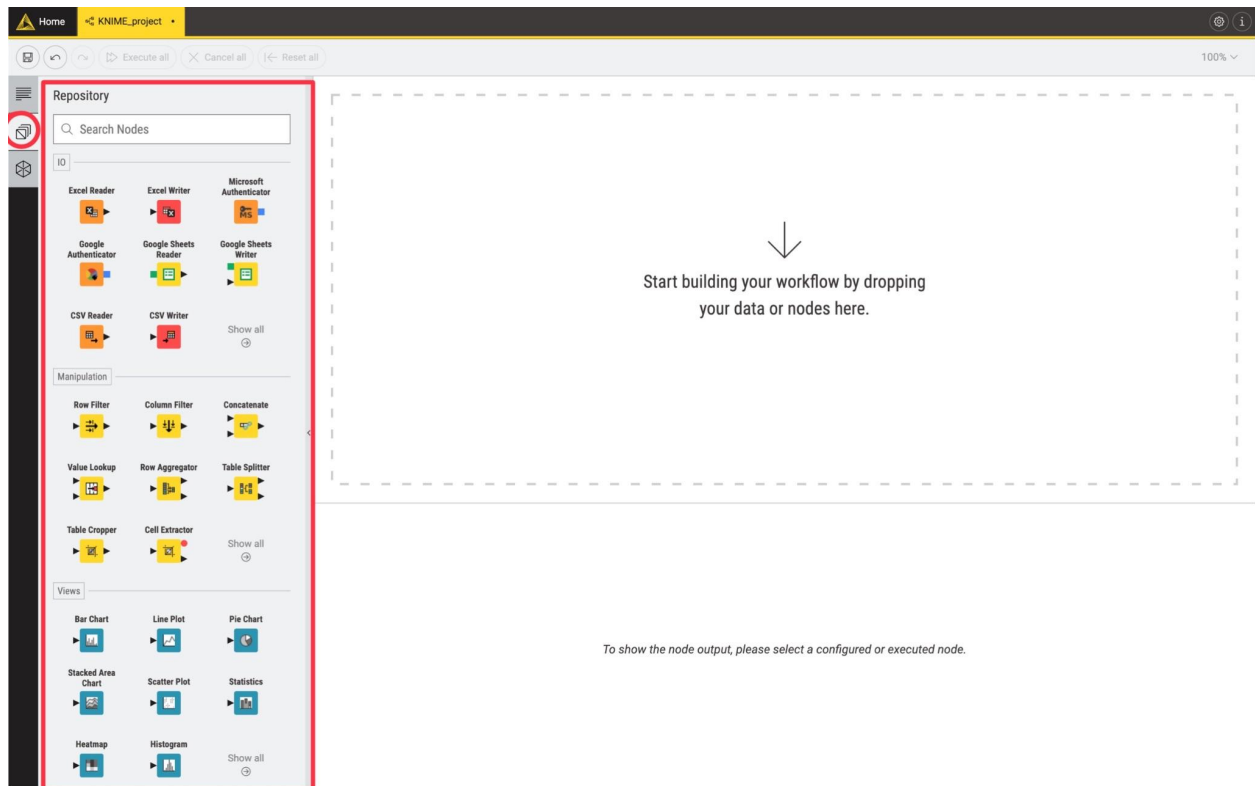


Example (Creating a simple workflow)

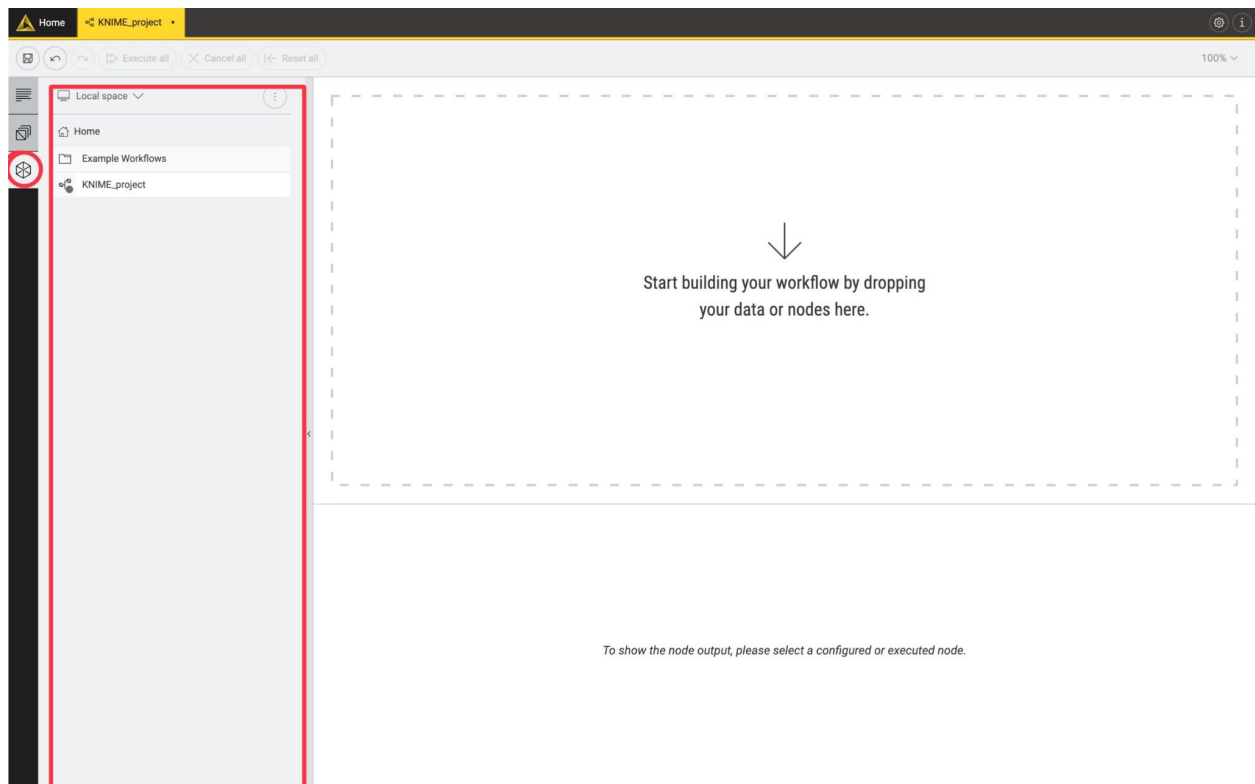
Step-by-step guide to read an Excel workbook and generate charts with it in KNIME.

Video link: <https://youtu.be/YPcNTHxnODM>

Repository - Has a list of all available and frequently used nodes



Directory - Has an image of the entire workspace



Description - Has a description of the selected node and its uses.

The screenshot shows the KNIME interface with the 'Bar Chart' node selected. The left sidebar displays the 'Bar Chart' configuration panel, which includes a description, input ports, and output image. The main workspace shows a workflow with an 'Excel Reader' node connected to a 'Bar Chart' node. Below the workflow, the 'Flow Variables' table is visible.

Bar Chart

The view can be accessed either via the "Open view" action on the executed node or on KNIME Hub.

In the node configuration, you can choose the aggregation method and enable certain controls, which are then available in the view. This includes the ability to choose different columns for the category and frequency axis or the possibility to set a title.

The configuration also offers a preview of the view, which should help to get the bar chart in the desired shape quickly.

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

Interactivity between multiple views is currently only possible for views coming from the [KNIME Views](#) extension.

Input ports

► **Type:**
Input Table
Data table containing the categories and values to be plotted in a bar chart.

Output Image (Dynamic Output)
Rendered PNG image.

Flow Variables

Count: 1

Owner ID	Data Type	Variable Name	Value
	StringType	knime.workspace	/Users/suduharidas/knime-workspace

Click and drag the required node into the workflow

The screenshot shows the KNIME interface with the 'Repository' panel open. The 'Excel Reader' node is highlighted with a red box, and a red arrow points from it to the main workspace. The workspace contains a large dashed box indicating where to drop nodes. Below the workspace, there is a text prompt: 'Start building your workflow by dropping your data or nodes here.'

Repository

Search Nodes

10

Excel Reader **Excel Writer** **Microsoft Authenticator**

Google Authenticator **Google Sheets Reader** **Google Sheets Writer**

CSV Reader **CSV Writer** **Show all**

Manipulation

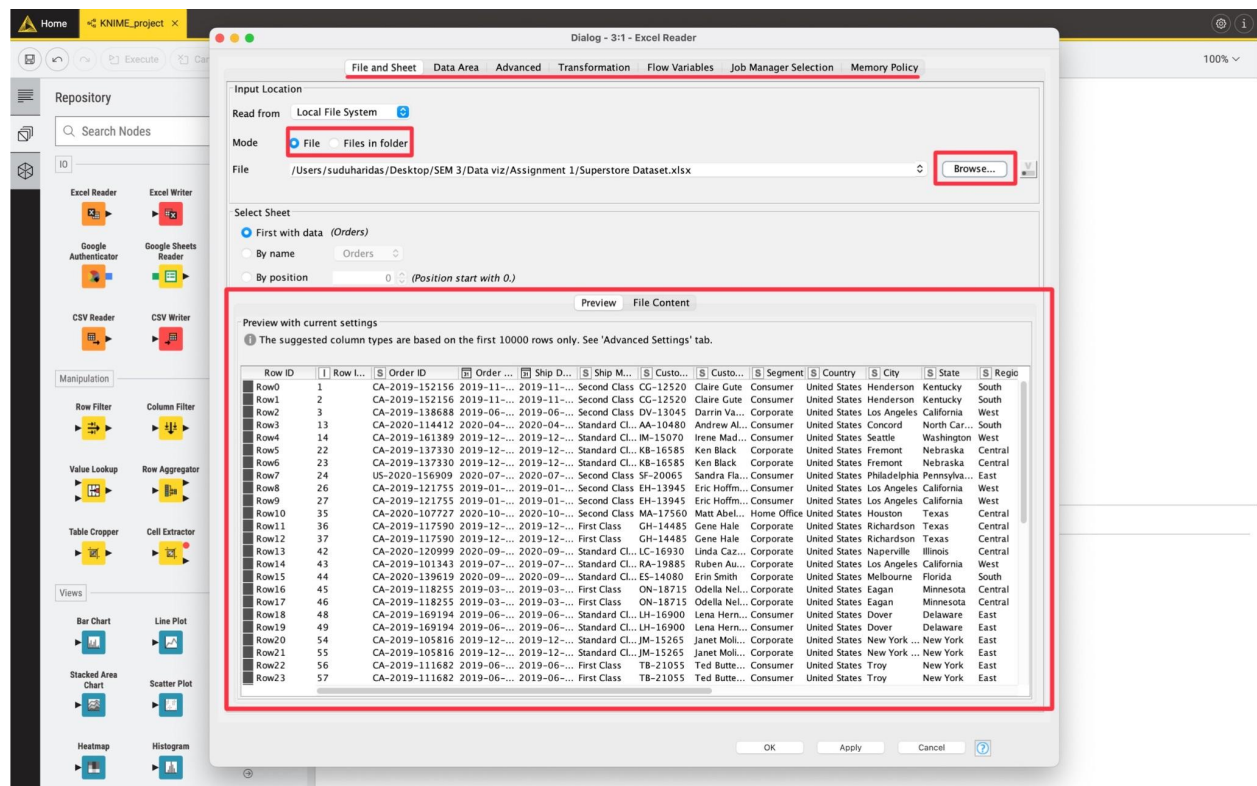
Row Filter **Column Filter** **Concatenate**

Value Lookup **Row Aggregator** **Table Splitter**

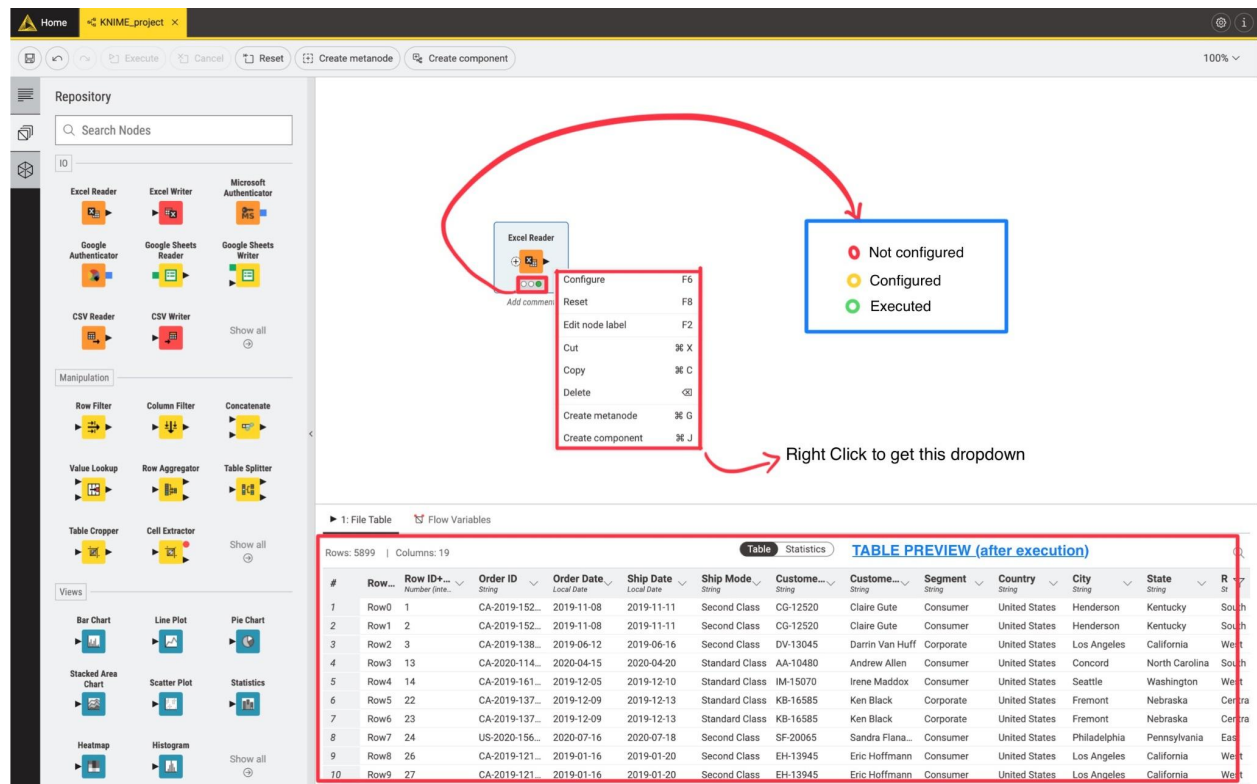
Table Cropper **Cell Extractor** **Show all**

Start building your workflow by dropping your data or nodes here.

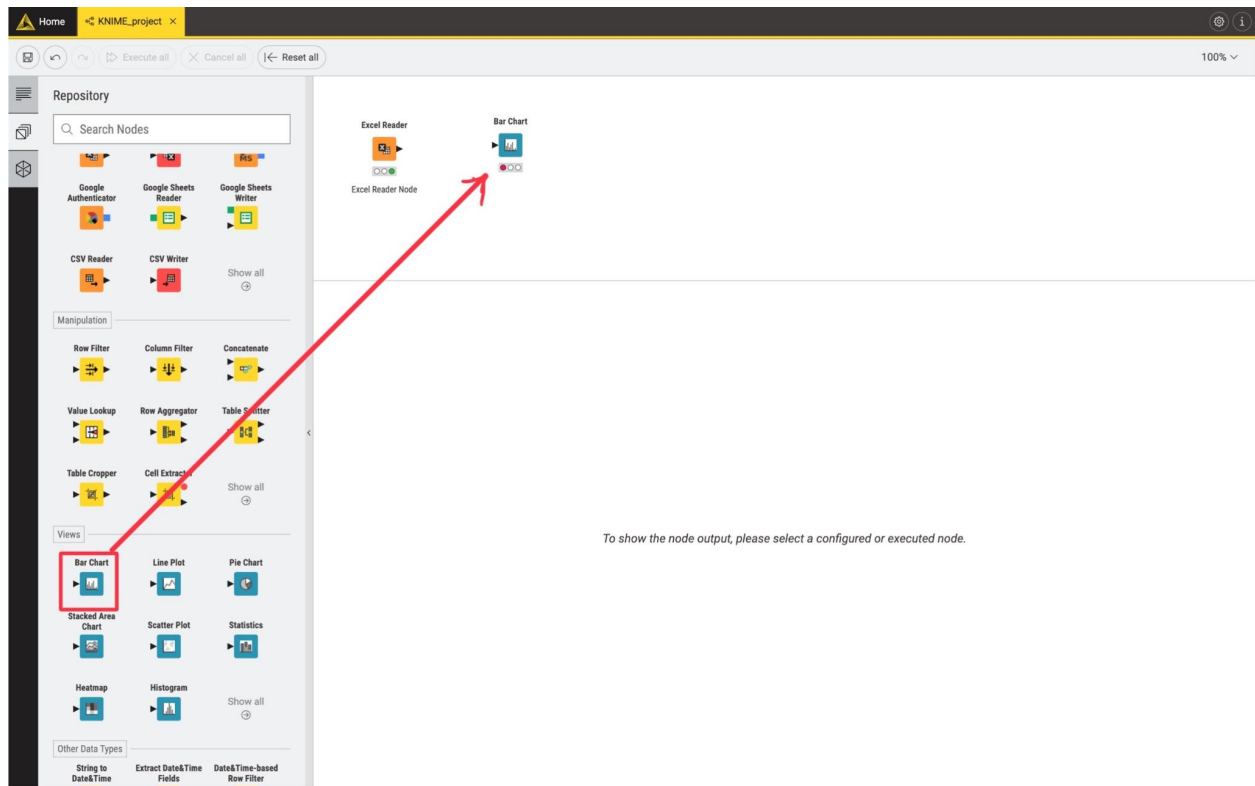
Right-click to configure the node and execute it.



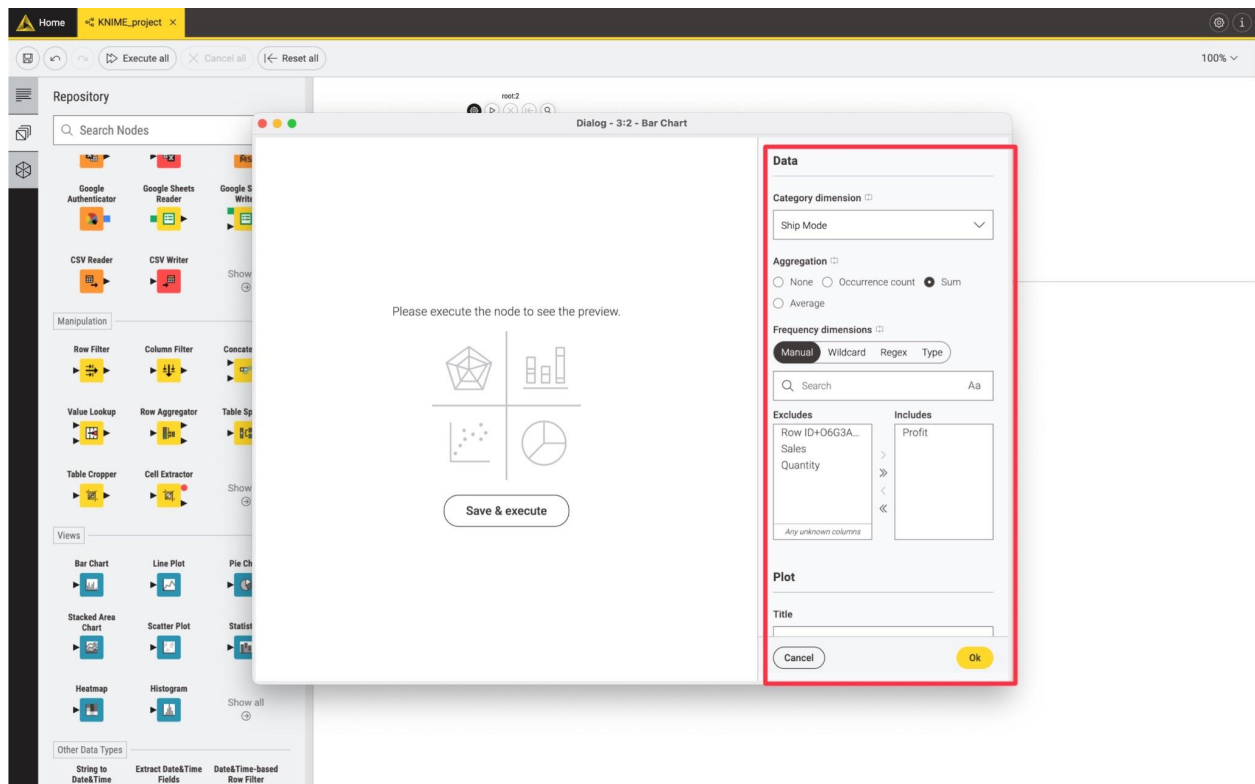
After execution, the console shows a preview of the completed action.



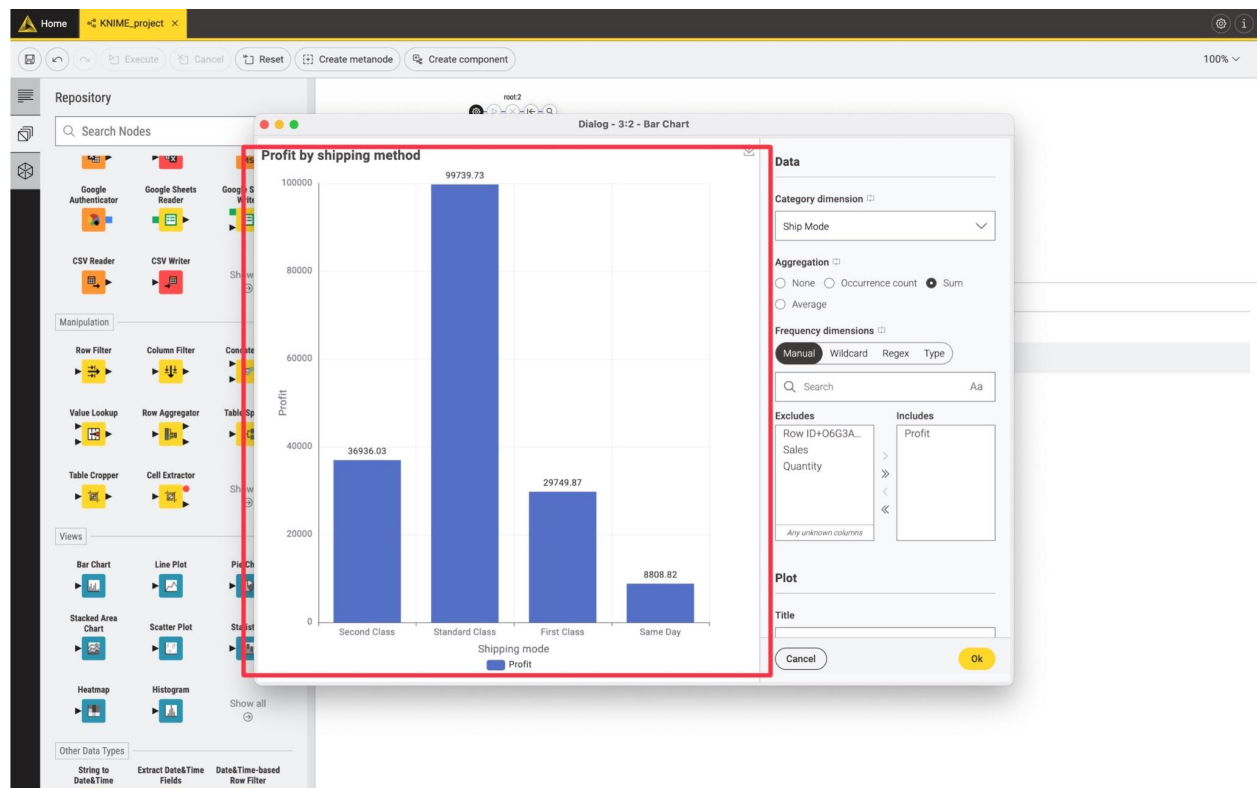
Drop the next required node into the workflow



Configure that node



Execute that node for preview



This is an example of creating a simple workflow in KNIME.

Conclusion

KNIME is a well-balanced open-source tool designed to create data workflows. Although KNIME lacks certain intricate capabilities, they can be custom-built using scripting languages such as Python. It also offers a vast community of data users to share their knowledge and collaborate.