Cheat Sheet: KNIME for Finance

Open for Innovation KNIME

MYDRAL

DATA ACCESS

Files

As for all reader nodes, you can add a "File system connection" input port to connect to external file systems like Amazon S3, Azure Blob Storage, etc.

Excel Reader: Reads content from sheets in Excel files (xlsx, xlsm, xlsb, and xls format). It can read single or multiple files, however, reading only one sheet per file.

CSV Reader: Reads content from CSV files. It has an auto-detect function to automatically guess the file structure.

Google Sheets Reader: Reads data from a Google Sheets spreadsheet after authenticating with the Google Authenticator node.

PDF Parser: Parses textual content and metadata from PDF files and creates a document for each file.

Tika Parser: Parses textual content and metadata and extracts embedded files and attachments from more than 280 file formats. Also provides an authentication option for encrypted files.

Databases

Dedicated connector nodes to connect to specific SQL, noSQL, or big data platforms, as well as to connect to data warehouses on the cloud. Only limited number of settings are required, e.g., hostname and credentials, and the necessary JDBC driver is already included.
The general **DB Connector** node can connect to any JDBC source, but it requires you to upload an appropriate drive and provide the JDBC URL.

SQL

noSQL

Big Data

Cloud

Data Lakes

Azure Data Lake Storage Gen2 Connector: Connects to Azure Data Lake Storage Gen2 (ADLS Gen2) and allows downstream nodes to access the ADLS Gen2 data to read or write files and folders or to perform other file system operations.

Amazon S3 Connector: Connects to Amazon S3 and points to a working directory (with a UNIX-like syntax, e.g., /mybucket/my-folder/myfile). Allows downstream reader nodes to access data from Amazon S3 as a file system.

Google Cloud Storage Connector: Connects to Google Cloud Storage and allows downstream nodes to access the Google Cloud Storage data of a project to read or write files and folders or to perform other file system operations.

Azure Blob Storage Connector: Connects to Azure Blob Storage and allows downstream nodes to access the Azure Blob Storage data as a file system.

Google Drive Connector: Connects to Google Drive and allows downstream nodes to access the files in Google Drive to read or write or to perform other file system operations.

Box Connector: Connects to Box and allows downstream nodes to access *files* to read or write or to perform other file system operations.

Misc

Sharepoint Online Connector: Connects to a SharePoint Online site and allows downstream nodes to access the document libraries to read or write files and folders or to perform other file system operations. The connection is closed when the node is reset, or the workflow is closed.

SAP Reader (Theobald Software): Accesses and loads data from various SAP systems (e.g., SAP S/4HANA, SAP BW, SAP R/3) via the Theobald Xtract Universal Server.

KCS SAP Executor: Connects to a wide range of SAP sources including SAP Tables, SAP Transactions (T-Codes) and Reports, SAP Spool, etc. It also accepts dynamic input filters.

Salesforce Simple Query: Reads fields from a Salesforce object into a KNIME table. It allows selecting the object type (i.e., a table in Salesforce such as *Account*) and the corresponding object fields (i.e., a column such as *Account Name*).

Google Sheets Connector: Connects to Google Sheets. Depending on the authentication method, the sheet should be either opened with a Google account or shared with a service account.

Google Analytics Connector: Connects to Google Analytics API.

GET Request: Issues HTTP GET requests to retrieve data from a web service without sending any data other than (optional) request parameters.

POST Request: Issues HTTP POST requests to send data to a web service and possibly receive data back.

TRANSFORMATION

GroupBy: Groups the rows of a table by the unique values in selected columns and calculates aggregation and statistical measures for the defined groups. It offers powerful functionality and has many unsuspected usages, for example, row deduplication.

Pivot: Extends the aggregation functionality of the **GroupBy** node by creating an output data table with columns and rows for the unique values in selected input columns. The unique values of the grouping column become rows and the unique values of the pivoting column become columns.

Joiner: Joins rows from two data tables based on common values in one or more key columns. The output - inner join, left outer join, right outer join, full outer join, or the respective antijoins - can be split into multiple output tables.

Concatenate: Merges two or more data tables vertically by piling up cells in columns with the same name. Cells in not overlapping columns are filled with missing values.

Row Filter: Filters rows in or out of the input table according to a filtering rule. The filtering rule can match a value in a selected column or numbers in a numerical range.

Column Filter: Filters columns in or out from the input data table according to a filtering rule. Columns to be retained can be manually picked or selected according to their type, or of a regex expression matching their name.

Rule Engine: Applies a set of rules to each row of the input data table. All operators are also available in the **Column Expressions** node.

String Manipulation: Performs operations on String values in columns, such as combining two or more Strings together, extracting one or more substrings, trimming blank spaces, and so on. All operators are also available in the **Column Expressions** node.

Math Formula: Implements a number of math operations across multiple input columns, from simple sum and average to logarithms and exponentials. All operators are also available in the **Column Expressions** node.

Cell Splitter: Splits values in a selected column into two or more substrings, as defined by a delimiter match. Delimiter is a set character, such as a comma, space, or any other character or character sequence.

Column Renamer: Assigns new names and types to selected columns as configured in the dialog.

Constant Value Column: Adds/replaces a column containing a single constant value in each row.

String to Date&Time: Converts values in a String column into Date&Time values. The Date&Time format contained in the String values can be manually defined or auto guessed.

Date&Time Shift: Shifts a selected date or time with a defined duration or granularity. The shift value can either be a duration column or a numerical column. A positive shift value is added to the selected date/time, a negative value will be subtracted.

Date&Time Difference: Calculates the difference between two Date&Time objects, e.g., from two selected columns, from a selected column and a fixed value, from a selected column and the current execution time, or from one cell and the cell in the previous row for a selected column.

Partitioning: Splits data into two subsets according to a sampling strategy. This node is generally used to produce a training and a test set to train and evaluate a machine learning model.

Sorter: Sorts the table in ascending or descending order based on the values of a chosen column. In addition, it is possible to sort based on multiple columns.

Missing Values: Defines a strategy to deal with missing values in the input data table - either globally on all columns, or individually for each single column.

WIDGETS

String Widget: Creates a text field input that outputs a String flow variable with a given value. Equivalent nodes exist for the creation of Integer, Double, Boolean, or Date&Time values.

Single Selection Widget: Allows selecting a single value from a set of values and outputs a String flow variable with the selected value. The set of possible values is available in the shape of menu, list, or radio buttons. Use the **Multiple Selection Widget** node for selecting multiple values.

Nominal Row Filter Widget: Creates a value filter widget that allows to interactively filter a data table in an Interactive View. The node takes a data table as input and outputs the filtered data table.

Column Selection Widget: Creates a column selection widget that allows to interactively filter a data table in an Interactive View. The node outputs a String flow variable with the name of the selected column(s).

Column Filter Widget: Creates a column filter widget that allows to interactively filter a data table in an Interactive View. Similar to the **Nominal Row Filter Widget** node, this node takes the data table as input and ouputs the filtered data table.

File Upload Widget: Creates a file upload item from which it is possible to navigate, select, and upload files. At the output port, it produces the file path as a variable.

File Download Widget: Provides a link with a downloadable file. The user needs to select a String or Path flow variable pointing to an existing file. This node is typically connected to a file writer (e.g. **CSV Writer node**), whereby the writer exposes its destination file as variable that is selected in this node's configuration dialog.

Date&Time Widget: Creates a calendar input item for date selection. It outputs a string flow variable with the selected value.

DATA VISUALIZATION

Bar Chart: Generates graphical representations of categorical data using rectangular bars, providing insights into category frequency or distribution.

Pie Chart: Visualizes categorical data using circular charts with slices representing categories and sizes indicating proportions.

Histogram: Displays the frequency distribution of a numeric variable, identifying patterns and anomalies.

Scatter Plot: Visualizes relationships between two numeric variables through points on a plane, identifying correlations, clusters, and patterns.

Sunburst Chart: Creates hierarchical visualizations representing the structure and composition of categorical or hierarchical data.

Radar Plot Appender: Appends radar plot attributes to data based on user-defined rules or mappings. Enables multivariate visualization and comparison.

Line Plot: Creates line charts to visualize trends, patterns, or correlations between two numeric variables.

Stacked Area Chart: Displays cumulative contributions of categories or variables, illustrating trends and relative proportions.

Statistics: Calculates descriptive statistics for selected numeric columns, aiding in data exploration and analysis.

Box Plot: Displays the distribution of a numeric variable with quartiles, median, and outliers. Useful for comparing distributions and identifying outliers.

Text View: Displays text output provided by a user. Useful to create text or number infographics in Interactive Views or Data Apps.

Tile View (Java Script): Displays tabular data in a grid layout, allowing for easy comparison and exploration of multiple data elements.

Table View: Allows viewing and inspecting data in a tabular format, facilitating exploration, sorting, and filtering.

Report PDF Writer: Writes a given report to a PDF file at the specified location.

Report Template Creator: Defines the basic layout of a report such as page layout and orientation. The template can then be passed to a component downstream, which will fill the report.

Resources

- **KNIME Press:** Access various data science books and other cheat sheets at ***knime.com/knimepress***, including beginner and advanced topics.
- **KNIME Blog**: Engaging topics, challenges, industry news, & knowledge nuggets at ***knime.com/blog.***
- **Self-Paced Courses:** Take our free online self-paced courses to learn about data analysis, data engineering, or data science with KNIME (with hands-on exercises) at ***knime.com/learning.***
- **KNIME Community Hub:** Browse and share workflows, nodes, and components or access collection pages for dedicated topics at ***hub.knime.com.***
- **KNIME Forum**: Join our global community & engage in conversations at ***forum.knime.com.***
- **KNIME Business Hub**: For team-based collaboration, automation, management, & deployment check out KNIME Business Hub at ***knime.com/knime-business-hub.***

VERIFIED COMPONENTS BY MYDRAL

Net Present Value (NPV): Computes the Net Present Value (NPV) just like in common spreadsheet tools. NPV tracks the total value of one or more investment projects with periodic transactions/cash flows.

Extended Net Present Value (XNPV): Computes the Extended Net Present Value (XNPV) just like in common spreadsheet tools. XNPV tracks the total value of one or more investment projects with non-periodic transactions/cash flows.

Check out the KNIME for Finance space on the KNIME Community Hub containing template solutions for:

- Accounting
- Audit & Compliance
- FP&A
- Financial Services
- Tax
- KPI

Internal Rate of Return (IRR): Calculates the Internal Rate of Return (IRR) just like in common spreadsheet tools. IRR tracks the profitability of one or more investment projects with periodic transactions/cash flows.

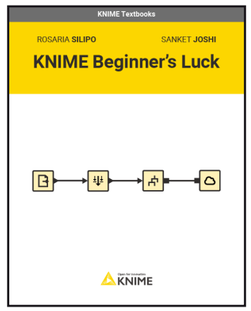
Modified Internal Rate of Return (MIRR): Computes the Modified Internal Rate of Return (MIRR) just like in common spreadsheet tools. MIRR analyzes the attractiveness of one or more investment projects with periodic transactions/cash flows and a reinvestment and finance rate.

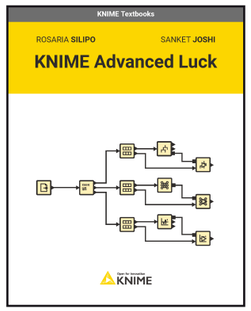
Extended Internal Rate of Return (XIRR): Computes the Extended Internal Rate of Return (XIRR) just like in common spreadsheet tools. XIRR tracks the profitability of one or more investment projects with non-periodic transactions/cash flows.

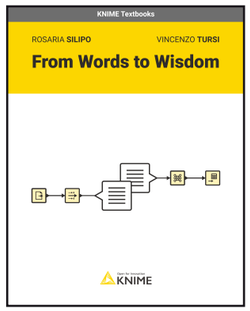
KNIME Press

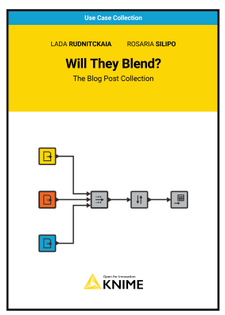
Open for Innovation
KNIME

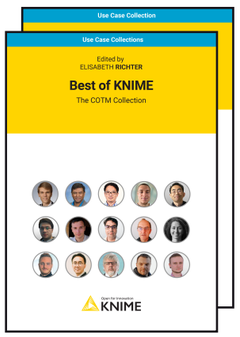
Extend your KNIME knowledge with our collection of books from KNIME Press. For beginner and advanced users, through to those interested in specialty topics such as topic detection, data blending, and classic solutions to common use cases using KNIME Analytics Platform - there's something for everyone. Available for download at www.knime.com/knimepress.

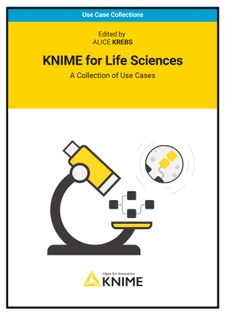


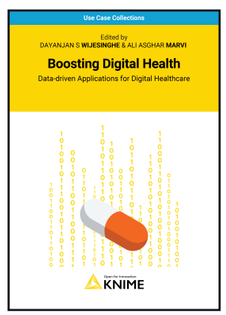


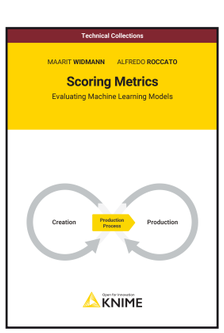


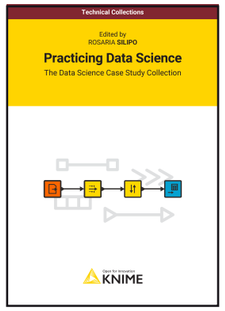


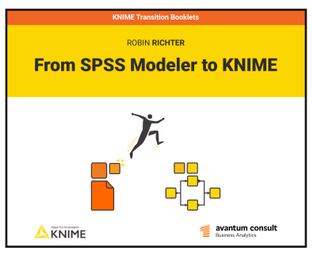


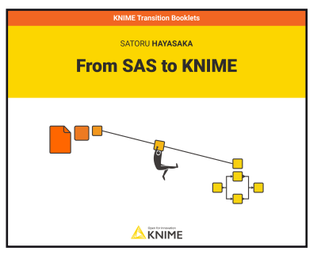


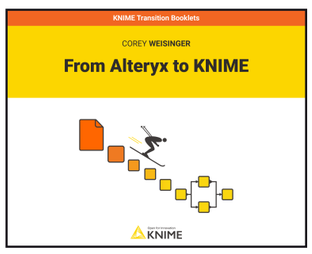












Need help?
Contact us!