

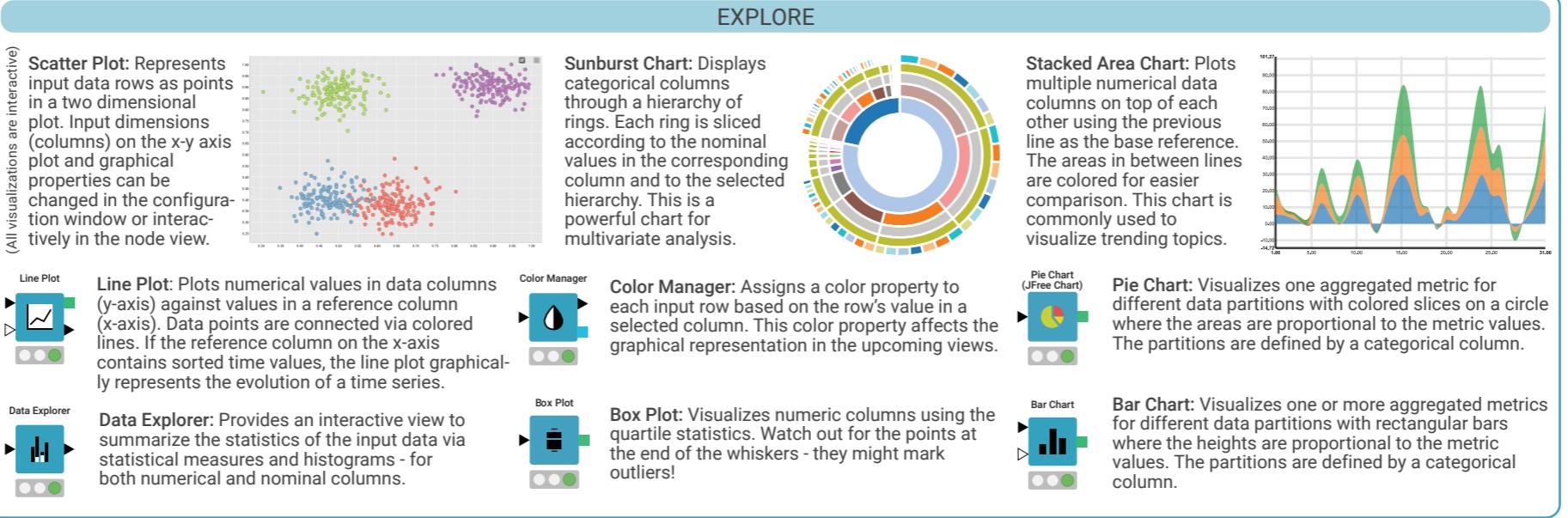
Cheat Sheet: Building a KNIME Workflow for Beginners

Getting started with KNIME Analytics Platform

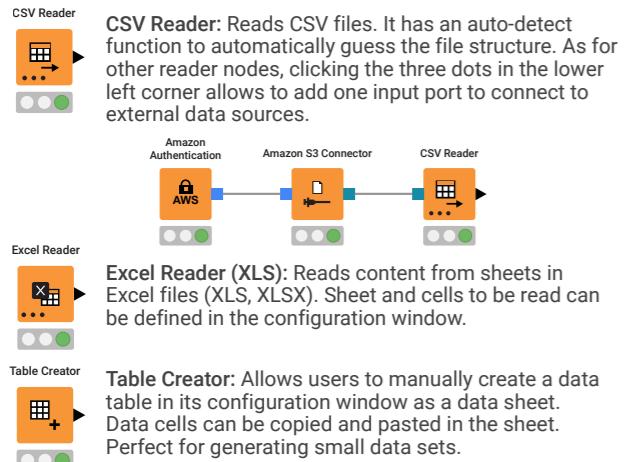
- Read through the installation guide at knime.com/downloads
- Check out the 7 things you should do after installing KNIME Analytics Platform at www.knime.com/blog/seven-things-to-do-after-installing-knime
- Take the E-Learning Course at www.knime.com/knime-self-paced-courses

Understanding the traffic light system:

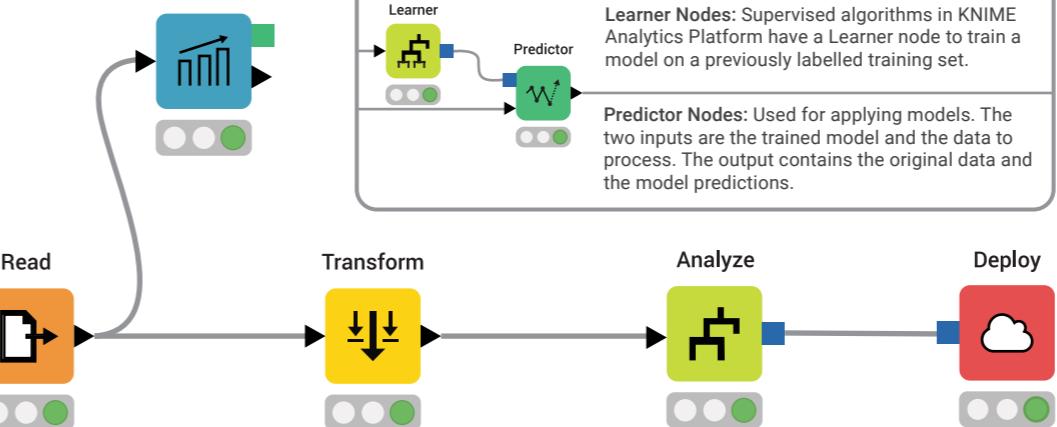
- Not configured: Node is not yet configured and cannot be executed with its current settings
 - Configured: Node has been correctly configured and may be executed at any time
 - Executed: Node has been successfully executed and results can be viewed and used in downstream nodes.
- Dynamic ports:** Additional input ports can be added by clicking the three dots in the bottom left corner of a node.



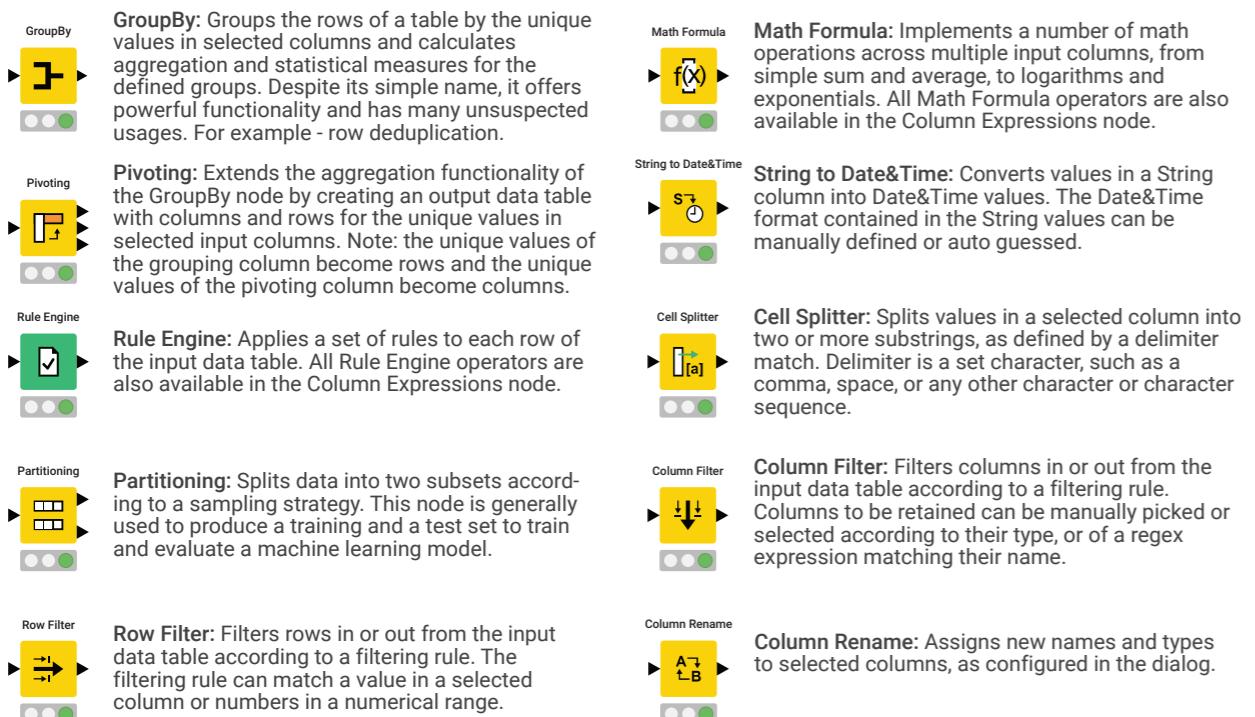
READ



Explore



TRANSFORM



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.

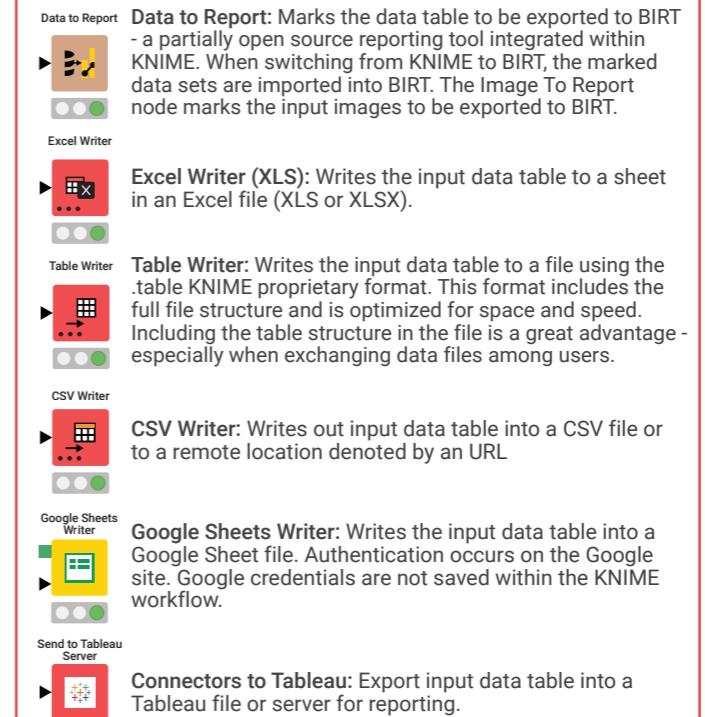
Sorter: Sorts the table in ascending or descending order based on the values of a chosen column. In addition, it is possible to sort based on multiple columns.

Concatenate: Merges two or more data tables vertically by piling up cells in columns with the same name. Cells in non-overlapping columns are filled with missing values.

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

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.

DEPLOY



ANALYZE

Decision Tree: The Learner node trains a C4.5 or a CART decision tree. The configuration window includes options for pruning, early stopping, information measures, splitting values, and more. Both the Learner and the Predictor node provide an interactive view where the decision tree is displayed together with the input data propagation.

k-Means: Implements the k-Means clustering algorithm. Number of clusters must be set prior to node execution. This node builds the clusters. The Cluster Assigner node finds the closest cluster and assigns it to the input data row. Being an unsupervised algorithm, this node pair doesn't follow the classic Learner - Predictor scheme.

Logistic Regression: The Learner node trains a logistic regression model to predict categorical target values. The configuration window includes options for solver, input feature choice, regularization functions to avoid overfitting, & more.

Scorer: Calculates a number of performance measures such as accuracy, F1-score, or Cohen's Kappa, to quantify the quality of a classifier.

Numeric Scorer: Calculates a number of numerical error measures, such as root mean squared error, mean absolute error, or R², to quantify the quality of a numerical predictor model.

ROC Curve: Displays the Receiver Operating Characteristic (ROC) curve of a classifier working on a binary class problem. One of the two classes is arbitrarily chosen as the positive class and the ROC curve is built on the probabilities/scores produced for that class on the input data set.

Integrations to many open source data analytics tools are also available. Some use the KNIME node GUI (H2O, Weka, Keras, Spark MLlib). Others offer nodes with a development environment for scripting and debugging (R, Python, Java).

Resources

- **KNIME Forum:** Join our global community and engage in conversations at forum.knime.com
- **KNIME Books:** More tips, ideas, and lessons from knime.com/knimepress
- **KNIME Events:** Take a course, attend a workshop, or join a meetup at knime.com/events
- **KNIME Blog:** Engaging topics, challenges, industry news, and knowledge at knime.com/blog
- **KNIME Hub:** Browse and share workflows, nodes, and components, with the KNIME community. Add ratings, or comments to other workflows at hub.knime.com
- **More Guides:** Still using SAS or Excel? Transition to KNIME Analytics Platform with these handy guides at knime.com/knimepress
- **KNIME Server:** For team-based collaboration, automation, management, and deployment check out KNIME Server at knime.com/knime-server
- **Beginners Space on KNIME Hub:** Find the collection of example workflows using these cheat sheet nodes.



KNIME Press

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.

