KNIME Textbooks

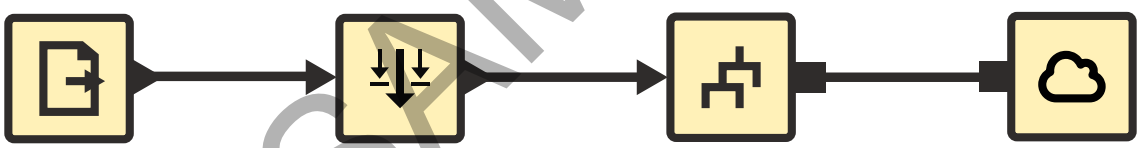
ROSARIA **SILIPO**

SANKET **JOSHI**

# KNIME Beginner's Luck

KNIME v5.2





Open for Innovation
KNIME

This book has been updated for **KNIME 5.2**.

For information regarding permissions and sales, write to:

KNIME Press
Talacker 50
8001 Zurich
Switzerland

knimepress@knime.com

ISBN: 978-3-9523926-5-2

www.knime.com

# Acknowledgements

First of all, I would like to thank the whole KNIME Team for their patience in dealing with me and my infinite questions.

Among all others in the KNIME Team I would like to specifically thank Peter Ohl for having reviewed this book in order to find any possible aspects that were not compatible with KNIME best practice.

I would also like to thank Casiana Rimbu for the help in providing the most beautiful, clear, and artistic screenshots I could ever imagine and Meta Brown for encouraging me in the first steps of developing the embryonic idea of writing this book.

Many thanks finally go to Heather Fyson for reviewing the book's English.

# Preface

This is the first book I wrote in 2010 for the KNIME Press on how to use KNIME Analytics Platform. Since then, we (the KNIME Press Team and I) have been constantly updating the book twice a year every year, following each new release of KNIE Analytics Platform; not immediately after – but close enough.

That is right! KNIME Beginner's Luck, like all other e-books from KNIME Press, is a live e-book, constantly changing to fit the newest version of the software – which is also true for the new and improved UX/UI that came with KNIME Analytics Platform Version 5 in summer 2023. This liveness of the e-book is also the reason why it has only rarely been printed. Updating printed pages is undoubtedly harder than updating a pdf file!

As this is the first book, it is inevitably about the basics: the basics of KNIME Analytics Platform of course and the basics of a data science project. This book guides you through the most important access functions, data transformation operations, and of course machine learning nodes available in KNIME Analytics Platform. Supplemented with many example workflows, exercises, and screenshots, it will quickly familiarize you with the basic functions of the software. If you are looking for more advanced topics, you won't find them here, instead….

If you want to learn more about advanced machine learning algorithms, flow variables, or loops, check the sequel to this book: "KNIME Advanced Luck". If you want to learn more about text processing, have a look at the book, "From Words To Wisdom". If you come from that school of thoughts where reading manuals or instructions is overrated, you can start directly with reading about solutions to case studies in various application fields in our collection "Practicing Data Science". If your job is more about integrating and blending different data sources and data types, then the book for you is the "Will they blend?" collection. More useful booklets are available on the KNIME Press page, if you are transitioning from Excel, Alteryx, SPSS Modeler, or SAS.

All this is to say that the KNIME Press team and I have been working hard to provide you with the learning material, books, and tutorials, to become progressively more and more productive with KNIME Software and data science concepts.

*Rosaria Silipo*

# Table of Contents

# Chapter 1: Introduction

## 1.1. Purpose and Structure of this Book

We live in the age of data! Every purchase we make is dutifully recorded; every money transaction is carefully registered; every web click ends up in a web click archive. Nowadays everything carries an RFID chip and can record data. We have data available like never before. What can we do with all this data? Can we make some sense out of it? Can we use it to learn something useful and profitable? We need a tool, a surgical knife that can empower us to cut deeper and deeper into our data, to look at it from many different perspectives, to represent its underlying structure.

Let's suppose then that we have this huge amount of data already available, waiting to be dissected. What are the options for a professional to enter the world of Business Intelligence (BI) and Data Science (DS)? The options available are of course multiple and growing rapidly. If our professional does not control an excessive budget, he or she could turn to the world of open-source software. Open-source software, however, is more than a money driven choice. In many cases it represents a software philosophy for resource sharing and control that many professionals support.

Inside the open-source software world, we can find a few Data Science and BI tools. KNIME Analytics Platform represents an easy choice for the non-initiated professional. It does not require learning a specific script and it offers a Graphical User Interface (GUI) to implement and document analysis procedures. In addition - and this is not a secondary advantage - KNIME Analytics Platform can work as an integration platform into which many other BI and Data Science tools can be plugged. It is then not only possible but even easy to analyze data with KNIME Analytics Platform and then to build dashboards on the same processed data with a different BI tool.

Even though KNIME Analytics Platform is very simple and intuitive to use, any beginner would profit from an accelerated orientation through all of the nodes, categories, and settings. This book represents the beginner's luck, because it is aimed to help any beginner to gear up his/her learning process. This book is not meant to be an exhaustive guide to the whole KNIME software. It does not cover implementations under the KNIME Business Hub, which is not open-source, or topics which are considered advanced. Flow Variables, for example, and

implementations of database SQL queries are discussed in the sequel book "KNIME Advanced Luck".

This book is divided into six chapters. The first chapter covers the basic concepts of KNIME Analytics Platform, while chapter two takes the reader by the hand into the implementation of the very first KNIME application. From the third chapter, we start the exploration of data science concepts in a more in-depth manner. The third chapter indeed explains how to perform some basic data exploration and visualization, in terms of nodes and processing flow. Chapter four is dedicated to data modeling. It covers a few demonstrative approaches to machine learning, Naïve Bayes, decision trees, and artificial neural networks. Finally, chapters five, six, and seven are dedicated to reporting. Usually, the results of an investigation based on data visualization or, in a later phase, on data modeling must be shown at some point to colleagues, management, directors, customers, or external workers. Thus, reporting is a very important phase at the end of the data analysis process. Chapter five shows how to prepare the data to export into a report, while chapter six shows how to build the report itself.

Each chapter guides the reader through an ETL or a machine learning (ML) process step by step. Each step is explained in detail and offers some explanations about alternative employments of the current nodes. At the end of each chapter several exercises are proposed to the reader to test and perfect what he/she has learned so far.

Examples and exercises in this book have been implemented using KNIME 5.2. They should also work under subsequent KNIME versions, although there might be slight differences in their appearance.



## 1.2. The KNIME Community

Being an open-source software, KNIME Analytics Platform benefits a number of forums and groups of KNIME users all around the world. This is a good safety net for advice, hints, and learning material. We report below the most popular sites and groups.

### Useful Web Pages

- *KNIME Website:* The root page on the KNIME website.

  https://www.knime.com

- *Software Overview:* The first place to look for an overview of all KNIME products. The open source KNIME Analytics Platform can be downloaded here.

  https://www.knime.com/software-overview

- *Learning Hub:* A central spot to access education material to get you started with KNIME.

  https://www.knime.com/learning

- *KNIME Community Hub:* The perfect place to search for nodes or example workflows when you're not quite sure what you need yet.

  https://hub.knime.com

- *KNIME Forum:* Come here to engage in community discussion, submit feature requests, ask for help, or help others yourself!

  https://forum.knime.com

- *Events and Courses:* Information on all our upcoming events including courses, webinars, learnathons, and summits.

  https://www.knime.com/events

- *Blogs:* A collection of blog posts covering data science with KNIME, a great space to learn what KNIME can really do.

  https://www.knime.com/blog

- *FAQ:* A collection of some of our most commonly asked questions, check out the forum if your answer isn't here!

  https://www.knime.com/faq

- *KNIME Press:* Information on all our available books, like this one!

  https://www.knime.com/knimepress

## Courses, Events, and Videos

- *Courses for KNIME Analytic Platform:* KNIME periodically offers onsite and online courses for the KNIME software. This includes basic and advanced elements. To check for the next available date and to register, just go to the KNIME Events web page (https://www.knime.com/events) and select the tab "Online Course".

- *KNIME Webinars:* A number of webinars are also frequently available on specific topics, like chemistry nodes, text mining, integration with other analytics tools, automated machine learning, deep learning, time series analysis, best practices, and so on. To know about the next scheduled webinars, check the KNIME Events web page (https://www.knime.com/events) and select the tab "Webinars".
- *KNIME Data Connects, KNIME Data Talks, and KNIME Summits:* KNIME Data Connects, KNIME Data Talks and KNIME Summits are held periodically all over the world. These are always good chances to learn more about the KNIME software, to get inspired about new data science projects, and to get to know other people from the KNIME Community. To check for the next upcoming events, just go to the KNIME Events web page (https://www.knime.com/events) and select the tabs "Data Talks", "Summit", or "Data Connect".
- *KNIME TV Channel on YouTube:* KNIME has its own video channel on YouTube, named KNIMETV. There, a number of videos are available to learn more about many different topics and specially to get updated about the new features in the new KNIME releases (http://www.youtube.com/user/KNIMETV).

## Books

- *Advanced Features in KNIME Analytics Platform:*

  For the advanced use: Rosaria Silipo & Victor Palacios, "KNIME Advanced Luck", KNIME Press

  https://www.knime.com/knimepress/advanced-luck

- *Data Science and KNIME*:

  For an overview of data science, data mining, and data analytics, please check: Berthold, M.R., Borgelt, C., Höppner, F., Klawonn, F., Silipo, R., "Guide to Intelligence Data Science", Springer 2020

  https://www.datascienceguide.org/

- *Codeless Deep Learning with KNIME*:

  It is possible to implement deep learning solutions also within KNIME Analytics Platform:

  Kathrin Melcher & Rosaria Silipo, "Codeless Deep Learning with KNIME", Packt, 2020

- *Codeless Time Series Analysis with KNIME*:

  A book explaining the main steps for time series analysis using the KNIME time series components.

  Corey Weisinger, Maarit Widmann, Daniele Tonini, "Codeless Time Series Analysis with KNIME", Packt, 2022

## KNIME Community Hub

However, there is a privileged place where to find information about KNIME nodes and example workflows for your next projects: the KNIME Community Hub (https://hub.knime.com/).

The KNIME Community Hub is a repository of applications, components, and nodes to recycle, reuse, and assemble on KNIME Analytics Platform. Or as it says on the home page: The KNIME Community Hub is "the place to find and collaborate on KNIME workflows and nodes. Here you can find solutions for your data science questions."

When you access the KNIME Community Hub the first time, you end up with the page in Figure 1.1. This page offers a few links to the starting guide documentation, the KNIME Forum, and

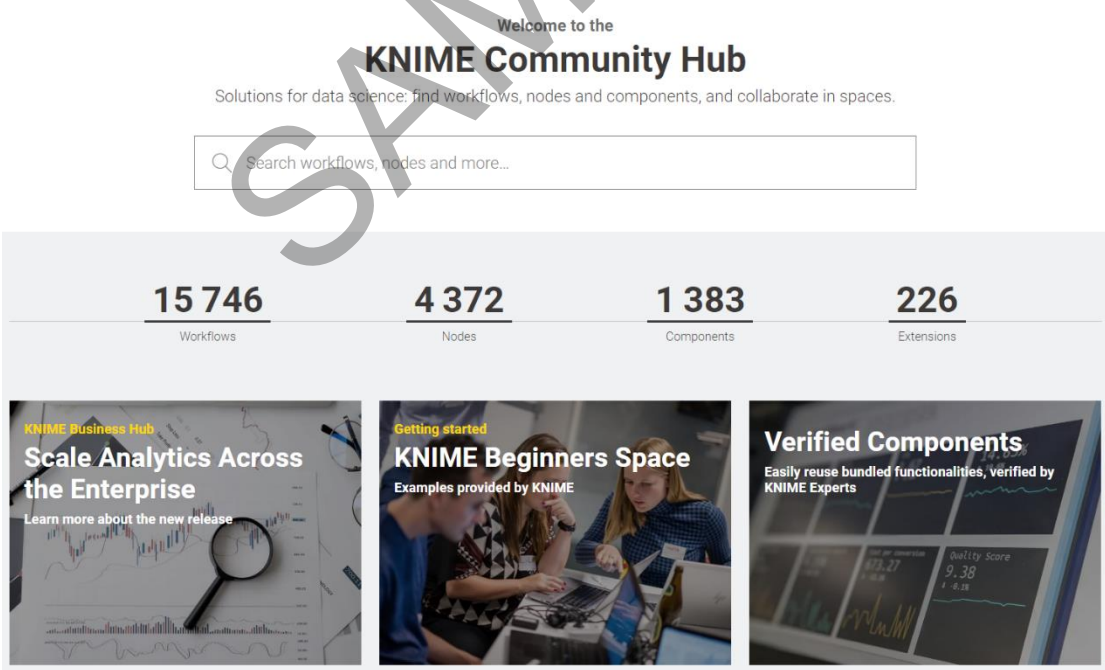



*Figure 1.1. The KNIME Community Hub home page at https://hub.knime.com.*

the KNIME blog. Most importantly at the top it offers a search box to find applications, nodes, and components uploaded by KNIME users in this shared place of the KNIME community.

If we type “Customer Intelligence” in the search box, we end up with a list of nodes and workflows related to Customer Intelligence. If you then select just “Workflows” in the top menu, you will see a list of applications (workflows) implementing some aspects of Customer Intelligence - and appropriately tagged - as uploaded by users of the KNIME community. Indeed, you can upload your own developed applications on the KNIME Community Hub. All you need is an account with the KNIME Forum.

On the KNIME Community Hub you can also find official dedicated spaces with their own description and landing page. Try for example to type “Marketing” and select the tag “Marketing Analytics”. You will get 31 marketing analytics related solutions hosted in the marketing analytics space and described on the landing page on the right, including relevant links.

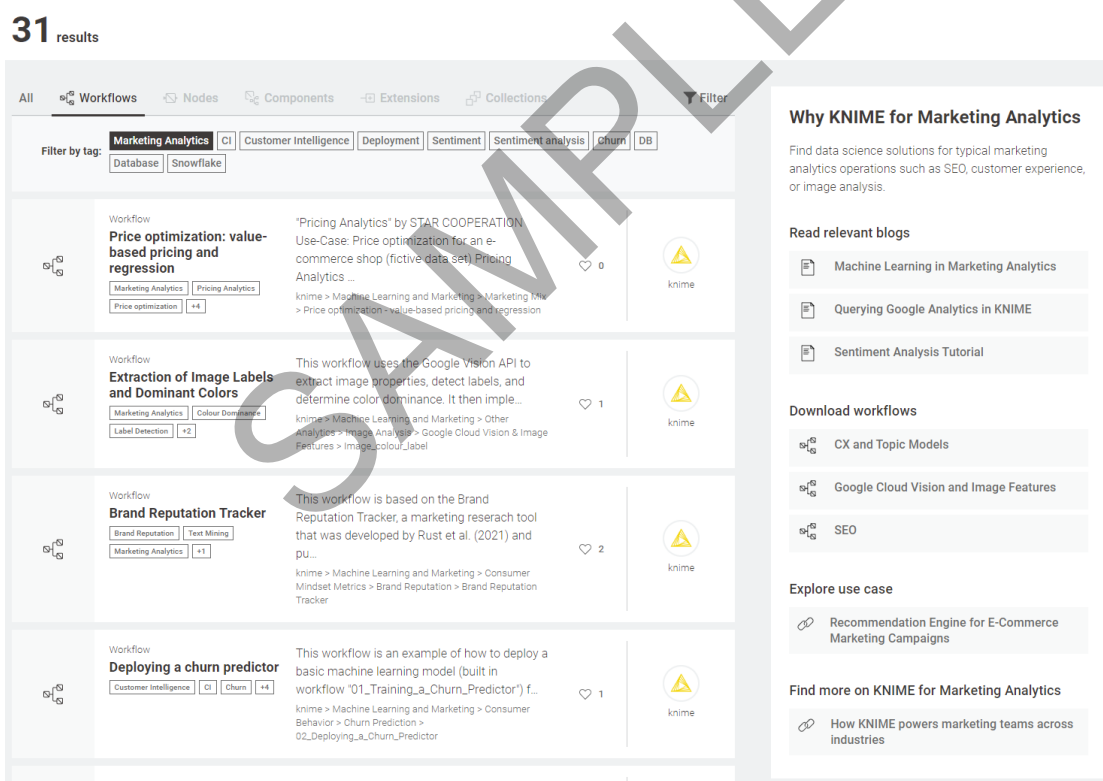



*Figure 1.2. The list of applications (workflows) related (and tagged) with “Marketing Analytics” on the KNIME Community Hub.*

Clicking one of the applications in the list opens the corresponding page (Figure 1.3), with a nice explanatory picture of the implemented workflow.

On the top right corner, you can see the button to log in with your KNIME account. Being logged in allows you to upload, download, comment, like, and update the spaces and workflows for which you have permissions. Below that, you can find the author picture and below that a number of utility buttons: to download the workflow, to like it, to drag & drop it into KNIME Analytics Platform, and to copy the short permanent link for this workflow to share.

If you hover over the author image, and if you have editing permissions for this Hub space, a pen will appear. Clicking on it will allow you to give other KNIME users permission to upload and change this space.

Notice that the KNIME Community Hub is a repository for workflows, but also for nodes, components, and extensions.

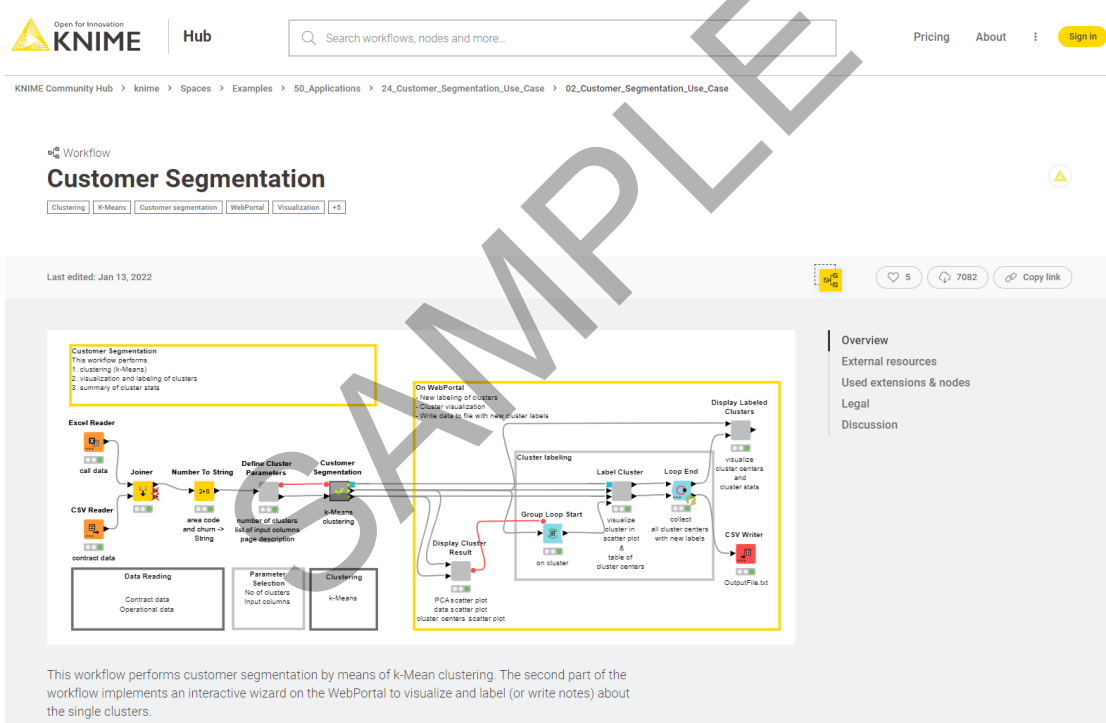





*Figure 1.3. The page dedicated to the application named "Customer Segmentation" on the KNIME Community Hub, with short link https://kni.me/w/37cHxqru6dbllUeP.*

## 1.3. Download and Install KNIME Analytics Platform

There are two available KNIME products:

- the open source KNIME Analytics Platform, which can be downloaded free of charge at https://www.knime.com/software-overview under the GPL version 3 license
- the KNIME Business Hub, which is described at https://www.knime.com/knime-business-hub

Analytically speaking, the functionalities of the two products are the same. The KNIME Business Hub in addition includes a number of useful IT features for team collaboration, enterprise workflow deployment and management, data warehousing, integration, and scalability for the data science lab. In this book, however, we will work with KNIME Analytics Platform (open source). To start playing with KNIME Analytics Platform, first, you need to download it to your machine.

## Download KNIME Analytics Platform

- Go to www.knime.com
- In the upper right corner of the main page, click “Download”
- Provide a little information about yourself (that is appreciated), then proceed to step 2 “Download KNIME”
- Choose the version that suits your environment (Windows/Mac/Linux, 32 bit/64 bit, with or without Installer for Windows) optionally including all free extensions
- Accept the terms and conditions
- Start downloading. You will end up with a zipped (*.zip), a self-extracting archive file (*.exe), or an Installer application
- For .zip and .exe files, just unpack it in the destination folder. If you selected the installer version, just run it and follow the installer instructions.

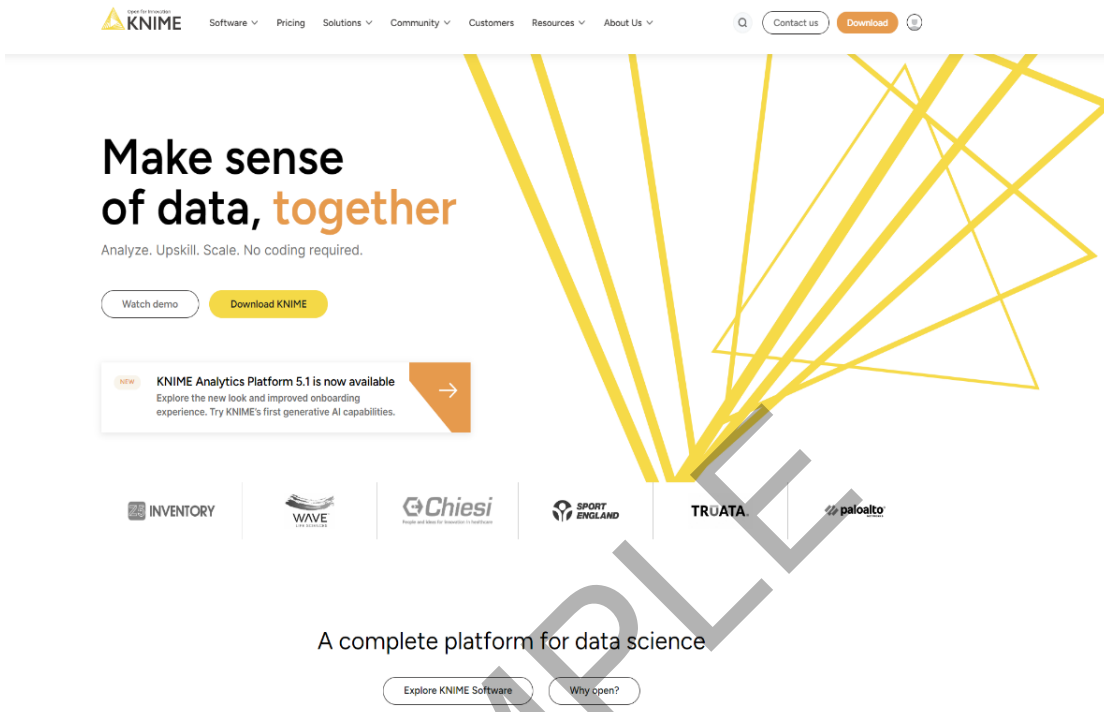



*Figure 1.4. The KNIME web page.*



## 1.4. Workspace

To start KNIME Analytics Platform, open the folder where KNIME has been installed and run knime.exe (or knime on a Linux/Mac machine). If you have installed KNIME using the Installer, then you can just click the icon on your desktop or on your Windows main menu.

If you are starting KNIME Analytics Platform for the first time, you will be asked if you want to share your usage statistics with KNIME. These numbers will be used to fuel the best practice recommendation engine provided within KNIME Analytics Platform workbench: the Workflow Coach. No personal information will ever reach KNIME and your anonymous statistics will never be shared with anybody.

After the splash screen, the “Workspace Launcher” window requires you to enter the path of the workspace.

## The Workspace Launcher

The *workspace* is a folder where all preferences and applications (workflows), both developed and currently under development, are saved for the next KNIME session.

The workspace folder can be located anywhere on the hard disk.

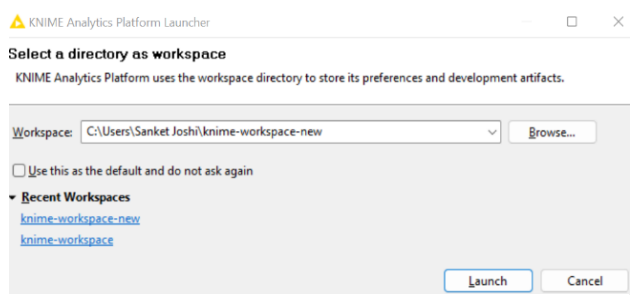



*Figure 1.5. The "Workspace Launcher" window.*

By default, the workspace folder is `"..\knime-workspace"`. However, you can easily change that, by changing the path proposed in the "Workspace Launcher" window, before starting the KNIME working session.

Once KNIME Analytics Platform has been opened, from within the KNIME workbench you can switch to another workspace folder, by selecting "File" in the top menu and then "Switch Workspace". After selecting the new workspace, KNIME Analytics Platform restarts, showing the workflow list from the newly selected workspace. Notice that if the workspace folder does not exist, it will be automatically created.

When having a large number of customers for example, different workspaces can be used for each one of them. This keeps each workspace clean and tidy and protects from mixing up information by mistake.

For this project I used the workspace "Knime-workspace-new".



# 1.5. KNIME Workflow

KNIME Analytics Platform does not work with scripts, it works with graphical workflows.

Small little icons, called nodes, are dedicated each to implement and execute a given task.

A sequence of nodes makes a workflow to process the data to reach the desired result.

## What is a Workflow?

A workflow is an *analysis flow*, i.e., the *sequence of analysis steps* necessary to reach a given result. It is the pipeline of the analysis process, something like:

*Step 1: Read data*
*Step 2: Clean data*
*Step 3: Filter data*
*Step 4: Train a model*

KNIME Analytics Platform implements its workflows *graphically*. Each step of the data analysis is implemented and executed through a little box, called *node*. A sequence of nodes makes a workflow.

In the KNIME whitepaper[1] a workflow is defined as follows: "Workflows in KNIME are graphs connecting nodes, or more formally, direct acyclic graphs (DAG)."

On the right is an example of a KNIME workflow, with:

- a node to read data from a file
- a node to exclude some data columns
- a node to filter out some data rows
- a node to write the processed data into a file

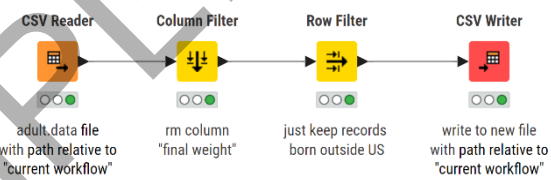



*Figure 1.6. Example of a KNIME Workflow.*



***Note.*** A workflow is a data analysis sequence, which in a traditional programming language would be implemented by a series of instructions and calls to functions. KNIME Analytics Platform implements it graphically. This graphical representation is more intuitive to use, lets you keep an overview of the analysis process, and makes for the documentation as well.

## What is a Node?

A node is the *single processing unit* of a workflow.

---

[1] *M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Koetter, T. Meinl, P. Ohl, C. Sieb, and B. Wiswedel, "KNIME: The Konstanz Information Miner". KDD 2006 (http://www.kdd2006.com/docs/KDD06_Demo_13_Knime.pdf).*

A node takes a data set as input, processes it, and makes it available at its output port. The "processing" action of a node ranges from modeling - like an Artificial Neural Network Learner node - to data manipulation - like transposing the input data matrix - from graphical tools - like a scatter plot, to reading/writing operations.

Every node in KNIME has 4 states:

- Inactive and not yet configured → **red** light
- Configured but not yet executed → **yellow** light
- Executed successfully → **green** light
- Executed with errors → **red with cross** light

Nodes containing other nodes are called *metanodes* or *components*.

On the right are four examples of the same node (a *File Reader* node) in each one of the four states.

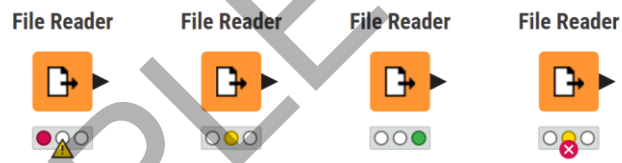



*Figure 1.7. The File Reader node in different states.*



## 1.6. File Extensions: *.knwf* and *.knar*

KNIME workflows can be packaged and exported in *.knwf* or *.knar* files. A *.knwf* file contains only one workflow, while a *.knar* file contains a group of workflows. Such extensions are associated with KNIME Analytics Platform. A double-click opens the workflow inside KNIME Analytics Platform.

| | | | |
|---|---|---|---|
| 01_From_Strings_to_Documents.knwf | 10/4/2017 9:45 AM | KNIME Workflow ... | 18,619 KB |
| 04_Interaction_Graph.knwf | 9/29/2017 8:20 AM | KNIME Workflow ... | 9,465 KB |
| 06_REST_Examples_Google_Geocode.knwf | 7/29/2017 7:09 PM | KNIME Workflow ... | 62 KB |
| 06_Semantic_Web_updated.knar | 11/3/2016 2:24 PM | KNIME Archive File | 178 KB |
| AzureDemoWorkflowArchive.knar | 5/5/2017 11:24 AM | KNIME Archive File | 24,104 KB |
| Building a Simple Classifier_.knwf | 2/18/2017 5:46 PM | KNIME Workflow ... | 43 KB |
| Cookbook_Ch5.knar | 11/24/2017 10:03 ... | KNIME Archive File | 477 KB |
| Cookbook_Ch6.knar | 11/24/2017 10:26 ... | KNIME Archive File | 155 KB |
| Corsair.knwf | 7/10/2017 4:20 PM | KNIME Workflow ... | 106 KB |

*Figure 1.8. .knwf and .knar files are associated with KNIME Analytics Platform. A double-click opens the workflow(s) directly inside the platform.*

# 1.7. KNIME User Interface

After accepting the workspace path, the KNIME UI opens on a “Getting started with KNIME Analytics Platform 5” entry page. This page provides access to the local space or create a new workflow. This page also provides the mount points. You can connect to them for usage by clicking the sign-in button. The “KNIME User Interface” consists of a top menu, a tool bar, and a few panels. Panels can be closed, re-opened.

- ***Home button:*** This panel shows the list of workflow projects available in the selected workspace (LOCAL), on the EXAMPLES, on the My-KNIME-Hub (your own space on the KNIME Community Hub), or on other connected KNIME Hub spaces.
- ***Quick node adding panel:*** This is a node recommendation engine. It will provide the list of the compatible nodes to follow the currently selected node. You will also be able to search for the other available nodes by typing in keywords in the search field. Double-click the node you want to add to the workflow, and it will be added and connected automatically.
- ***Node Repository:*** This panel contains all the nodes that are available in your KNIME installation. It is something similar to a palette of tools when working in a report or with a web designer software. There we use graphical tools, while in KNIME we use data analytics tools.

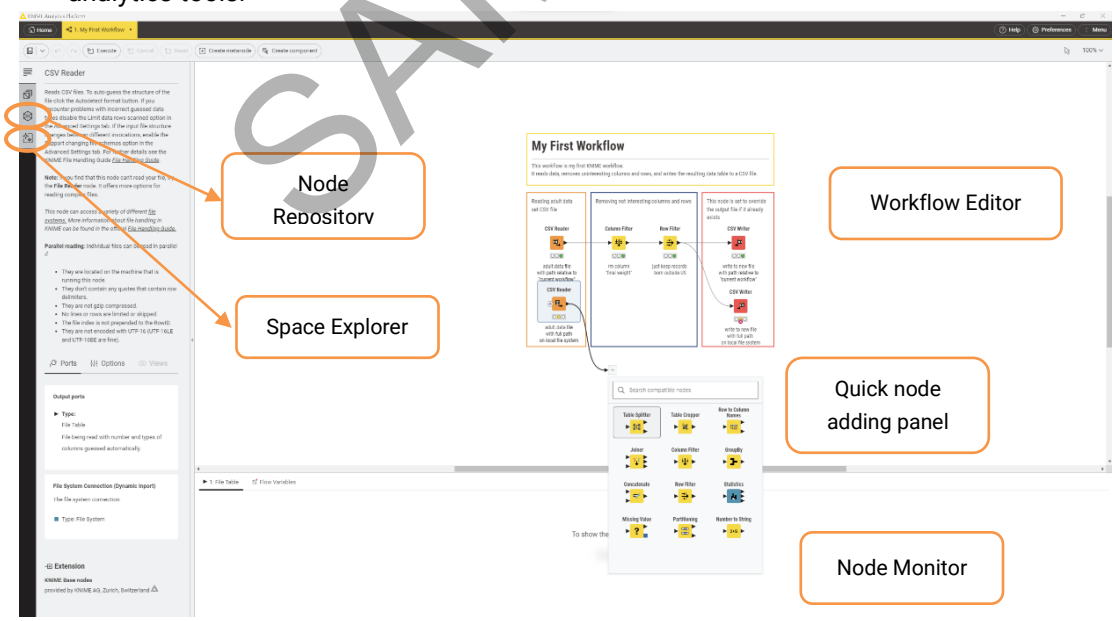



*Figure 1.9. The KNIME user Interface.*

- ***Workflow Editor:*** Workflow Editor is the central part of the KNIME UI. A node can be selected from the "Node Repository" panel and dragged and dropped here, in the "Workflow Editor" panel. Nodes can be connected by clicking the output port of one node and releasing the mouse either at the input port of the next node or at the next node itself.
- ***Space Explorer:*** To navigate to local or KNIME Hub spaces and access the workflows, components, and files, you can switch to the "Space Explorer" panel.
- ***Node Description:*** If a node or a workflow is selected, the "Node Description" panel on the left displays a summary description of the node's functionalities or the workflow's meta information.

## 1.8. Help

The "Help" button is one of the newer options added to the KNIME User Interface in version 5.2.

It includes a few more buttons; if you click on those buttons, you will be redirected to the respective website.

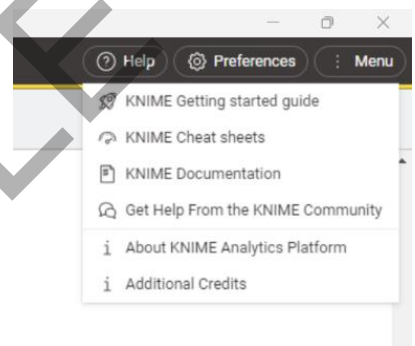



*Figure 1.10. The "Help" button in KNIME.*

## 1.9. Preferences

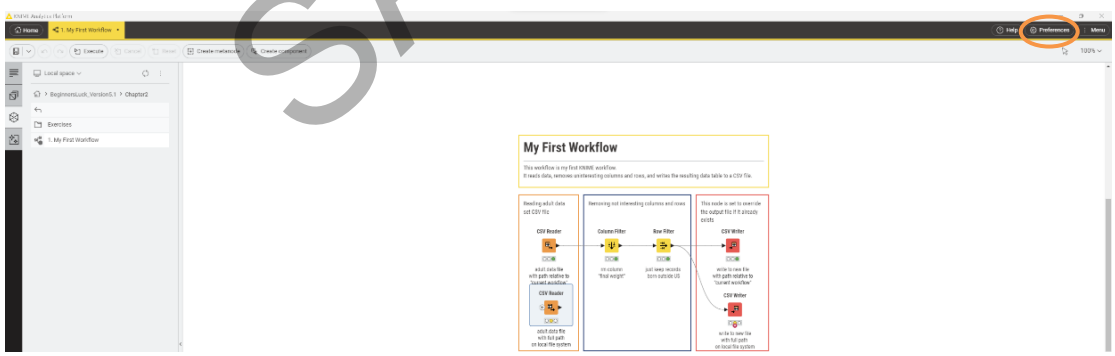



*Figure 1.11. The "Preferences" option in the KNIME User Interface.*

Preferences brings you to the window where all KNIME settings can be customized. Under the item "KNIME" can be found:

- ***Chemistry*** – has settings related to the KNIME Renderers in the chemistry packages.

- ***Databases*** – specifies the location of specific database drivers, not already available within KNIME. Indeed, the most common and most recent database drivers are already available in the driver menu of Database nodes. However, if you need some specific driver file, you can set its path here.
- ***Space Explorer*** – contains the list of the shared repositories via KNIME Hub spaces.
- ***KNIME GUI*** – allows the customization of the KNIME workbench options and layout via a number of settings.
- ***Master Key*** – contains the master key to be used in nodes with an encryption option, like database connection nodes. Since KNIME 2.3 database passwords are passed via the "Credentials" workflow variables and the Master Key preference has been deprecated. You can still find it in the Preferences menu for backward compatibility.
- In ***Meta Info Preferences*** you can upload meta-info template for nodes and workflows.
- Here you can also find the preference settings for the ***external packages***, like ***H2O, R, Report Designer, Perl, Perl, Open Street Map,*** and others if you have them installed. In particular, for the external scripts, this page offers the option to set the path to the reference script installation.
- Finally, ***Workflow Coach*** contains the dataset to be used for the node recommendation engine: the community, a Hub workspace, or your own local workspace.

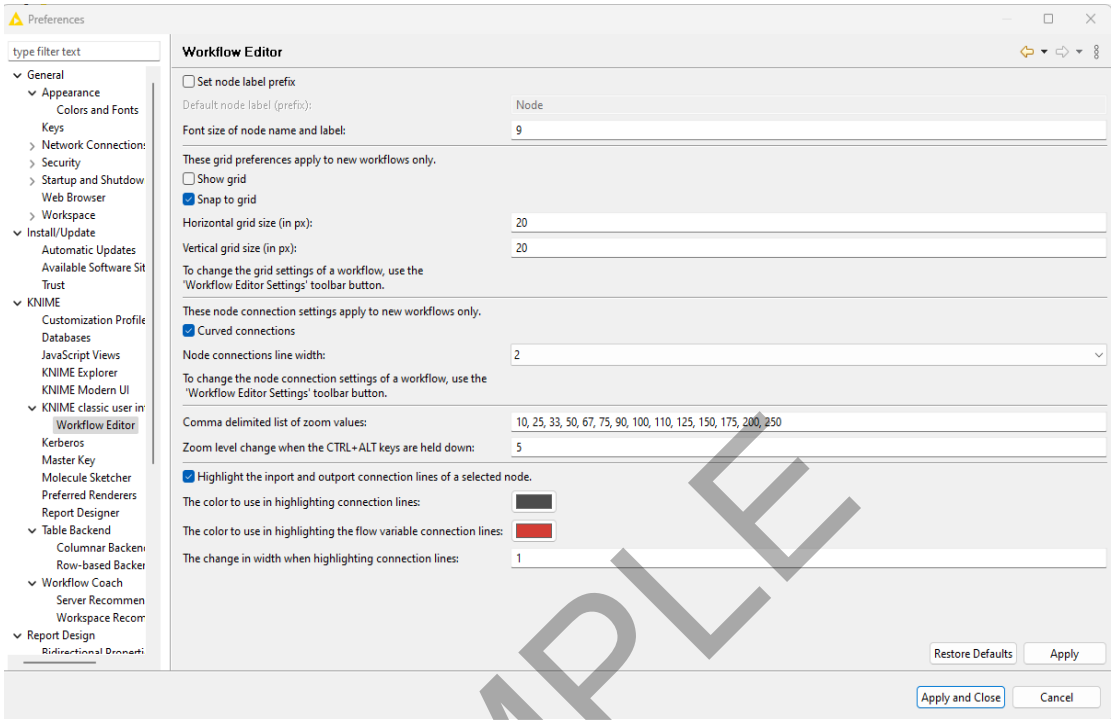



*Figure 1.12. The "Preferences" window.*

## Tool Bar

The tool bar is another important piece of the KNIME User Interface.

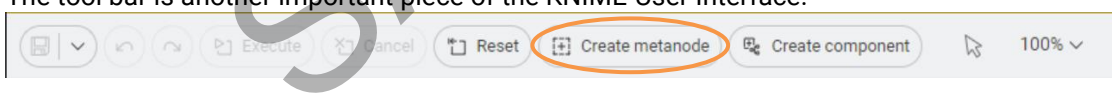



*Figure 1.13. The "Create Metanode" button in the tool bar.*

From the left, we see

- an icon to *save* or *save as* the current workflow. To the left, we find the icon to save the selected workflow
- *undo* and *redo* the changes that were made to the workflow
- *execute* or *execute all* (which executes all the selected nodes)
- *cancel* the executing node(s)
- *reset* all the node(s) to the original state
- *create Metanode*

- *create Component*,
- the arrow icon which provides the option of selecting, using the annotating mode or using the panning mode
- and on the right is the zoom (in %) level selection.

For now, let's have a look at the ***Create Metanode*** button. The *Create Metanode* button creates a metanode for all the selected nodes. It creates one node inside which you can find all the selected nodes. This is particularly useful to create a clean and easy-to-use workflow, for example, by compressing all the nodes used for data cleaning within a *Pre-processing* metanode.

## Hotkeys

For all keyboard lovers, most KNIME commands can also run via hotkeys. All hotkeys are listed in the KNIME menus on the side of the corresponding commands or in the tooltip messages of the icons in the Tool Bar under the Top Menu. Here are the most frequently used hotkeys.

**Node Configuration**

- `F6` opens the configuration window of the selected node

**Node Execution**

- `F7` executes selected configured nodes
- `Shift + F7` executes all configured nodes

**Stop Node Execution**

- `F9` cancels selected running nodes
- `Shift + F9` cancels all running nodes

**Node Resetting**

- `F8` resets selected nodes

**Save Workflows**

- `Ctrl + S` saves the workflow
- `Ctrl + Shift + S` saves all open workflows

- `Ctrl + W` closes all open workflows

**Metanode**

- `Shift + F12` opens Meta Node Wizard

**To move Annotations**

- `Ctrl + Shift + PgUp/PgDown` moves the selected annotation in the front or in the back of all the overlapping annotations

# 1.10. Menu

The “Menu” option button is another option added to the KNIME User Interface in KNIME Analytics Platform v5.2.

Once you click this button, you will see multiple options, and we will see what each option does.

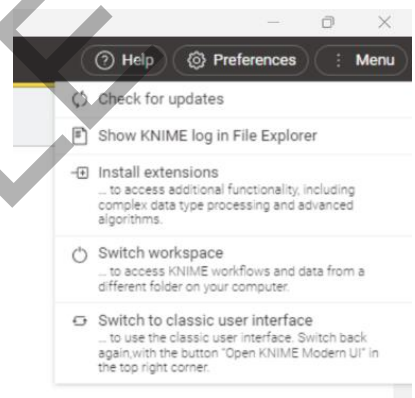



*Figure 1.14. The "Menu" button in the tool bar.*

- ***Check for Updates:*** On clicking this option, you can see if any updates are available for your KNIME Analytics Platform.
- ***Show KNIME log in File Explorer:*** This option will open a window in your local system where the KNIME log file is stored.
- ***Install Extensions:*** On clicking this option, you will be able to see a list of extensions that you could install.
- ***Switch Workspace:*** This will allow you to change the working directory.
- ***Switch to Classic User Interface:*** This option will first ask you to save or not save the current workflow and then switch to the classic UI. You can also switch back to the Modern UI from the classic UI by clicking on the “Open KNIME Modern UI” option in the top right corner of the Analytics Platform.

# 1.11. Node Repository

In the lower left corner, we find the Node Repository, containing all installed nodes organized in categories and subcategories. KNIME Analytics Platform has accumulated by now more than 1500 nodes. It has become hard to remember the location of each node in the Node Repository. To solve this problem, two search options are available: by exact match and by fuzzy match, both in the search box placed at the top of the Node Repository panel.

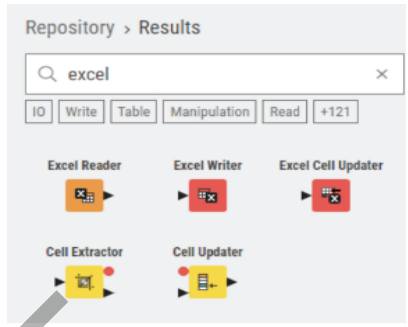



*Figure 1.15. Word search in the Node Repository panel: exact match mode.*

### Search Box

At the top of the “Node Repository” panel there is a search box. If you type a keyword in the search box and hit “Enter”, you obtain the list of nodes containing an exact match of that keyword. Press the “Esc” key to see all nodes again.

# 1.12. Space Explorer

We find the Space Explorer panel below the Home button, one of the four panels on the KNIME User Interface. This panel contains:

- Under LOCAL the workflows that have been developed in the selected workspace
- The mount points to a number of KNIME Hub spaces.
- The workflows contained in the reference workspace of such Hub spaces.
- The access to the My-KNIME-Hub, that is to your space on the KNIME Community Hub. Remember that you need an account with the KNIME Forum to access this space.

At the beginning, the Space Explorer panel only contains LOCAL, My-KNIME-Hub, and EXAMPLES. As we already stated, LOCAL shows the content of the selected workspace. EXAMPLES points to a read-only public repository, accessible via anonymous login. This repository hosts a number of example workflows that you can use to jump start a new project. My-KNIME-Hub allows to access your space on the KNIME Community Hub.

When you open KNIME Analytics Platform for the first time, you will find a folder named "Example Workflows" containing the solutions to a few common data science use cases, comprehensive of data.

Folders in "Space Explorer", containing workflows, are also called "Workflow Groups".

***Note.*** Space Explorer panel can also host data. Just create a folder under the workspace folder, fill it with data files on the machine, and select "Refresh" in the context-menu (right-click) of the "Space Explorer" panel.

## My-KNIME-Hub

From the Space Explorer panel, you can access your spaces on the KNIME Community Hub and upload and update new or existing content in there.

By default, an authenticated KNIME user has a public space, for material to share publicly, and a private space to park his/her own material for further usage. However, new private or public spaces can be created with a right-click on My-KNIME-Hub in the Space Explorer panel and then a selection of the option "Create new Space…".

By default, you are the only owner of your own spaces. However, when accessing this space from a web browser, after hovering on your image in the top right corner, a pen appears. This will allow us to add colleagues and teammates as contributors to the space.

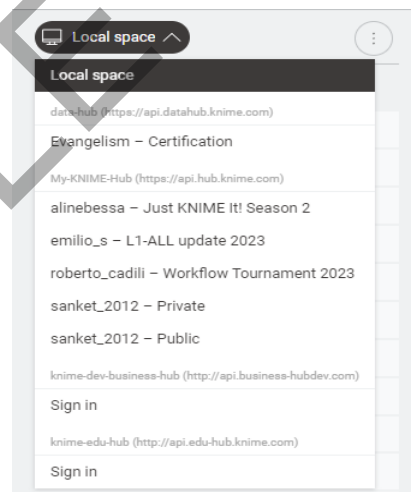



*Figure 1.16. Space Explorer panel. At the top the content of the EXAMPLES; below the content of the LOCAL workspace.*

## EXAMPLES

A link to EXAMPLES is available in the "Space Explorer" panel. This is a repository provided by KNIME to all users for tutorials and demos. There you can find a number of useful examples on how to implement specific tasks with KNIME. To connect to the EXAMPLES:

- double click "EXAMPLES" in the "Space Explorer" panel
- double click "Double click to connect…"

You should be automatically logged in as a guest.

To transfer example workflows from the EXAMPLES to your LOCAL workspace, just drag and drop or copy and paste (Ctrl-C, Ctrl-V in Windows) them from “EXAMPLES” to “LOCAL”.

You can also open the EXAMPLES workflows in the workflow editor, however only temporarily and in read-only mode. A yellow warning box on top warns that this workflow copy will not be saved.

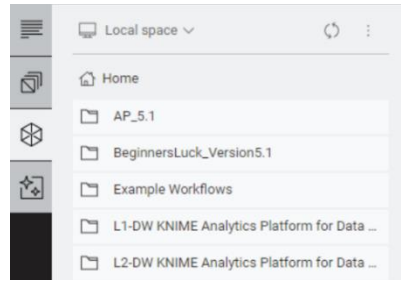



*Figure 1.17. Space Explorer panel.*

The Space Explorer panel can of course host more than one KNIME Hub space. It is enough to add mountpoints to the list of the available KNIME Hub spaces.

## Mounting KNIME Business Hub in Space Explorer

To add KNIME Business Hubs to the “Space Explorer” panel, in the modern UI, click on “Preferences” -> KNIME -> KNIME Explorer and:

- The “Preferences (Filtered)” window opens on the “KNIME Explorer” page and lists all KNIME spaces already mounted in this KNIME instance. The three KNIME spaces available by default on every KNIME instance are the local workspace “LOCAL”, the KNIME “EXAMPLES”, and the My-KNIME-Hub located on the KNIME Community Hub (hub.knime.com).
- Use the “New” and the “Remove” button to add /remove connections to remote Hub spaces.
- After clicking the “New” button, fill in the required information about the KNIME Business Hub in the “Select New Content” window (Figure 1.18).

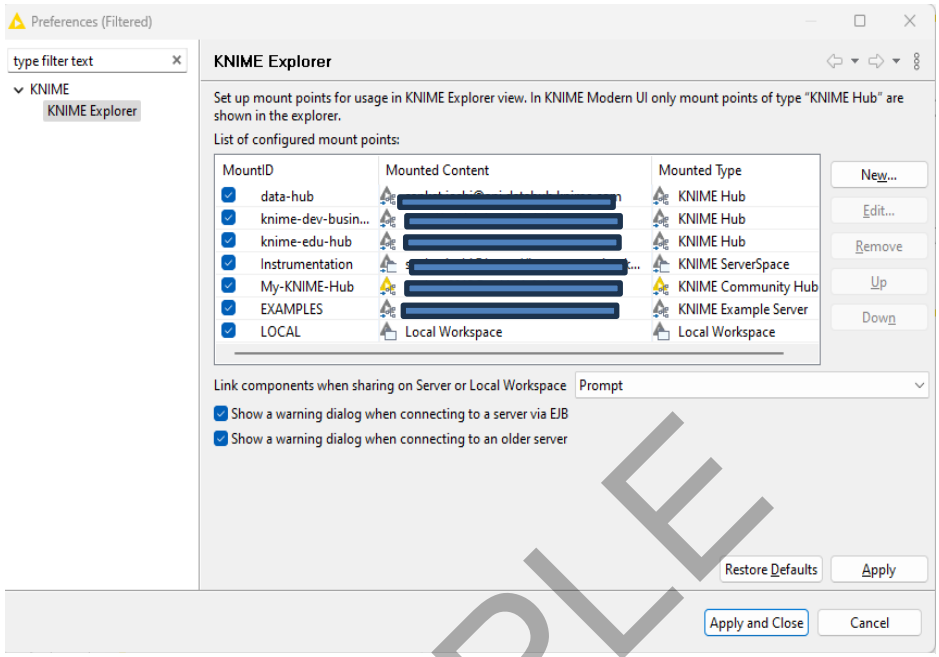



*Figure 1.18. The "Preferences (Filtered)" window.*

The same KNIME Explorer "Preferences" page can be reached via *File > Preferences > KNIME Explorer*.

To login into any of the available KNIME Business Hubs in the "Space Explorer" panel:

- right-click or double-click the Hub space name
- provide the credentials

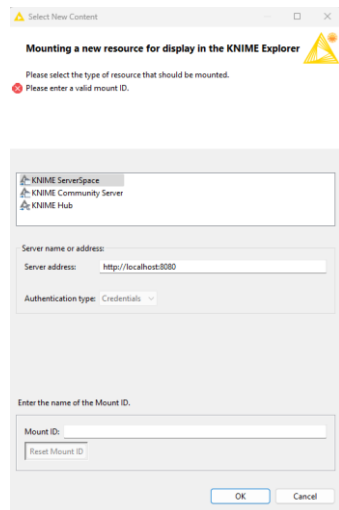



*Figure 1.19. The "Select New Content" window.*

## Workflow Editor

The central piece of the KNIME User Interface consists of the workflow editor itself. This is the place where a workflow is built by adding one node after the other. Nodes are inserted in the workflow editor by drag and drop or double-click from the Node Repository or the Quick node adding panel. The workflow building process will be described widely in the next sections of this book. Here, we will describe how to customize and probably improve the canvas role of the workflow editor space. We will describe two options:

- change the canvas appearance with grids and different visualizations for the connections;
- introducing annotations to comment the work.

### Adding annotations to the canvas

It is also possible to include annotations in the workflow editor. Annotations can help to explain the task of the workflow and the function of each node or group of nodes. The result is an improved documentation-like overview of the workflow general task and of the single sub-tasks.

**Workflow Annotations**

To insert a new annotation:

- right-click anywhere in the workflow editor and select “New Workflow Annotation”
- a gray small frame appears; this is the default annotation frame
- double-click the frame to edit its content
- Notice the tool bar appearing at the top to edit text style, text size, background color, text alignment, and border properties (color, thickness)
- To reopen an annotation, just double-click anywhere on the annotation

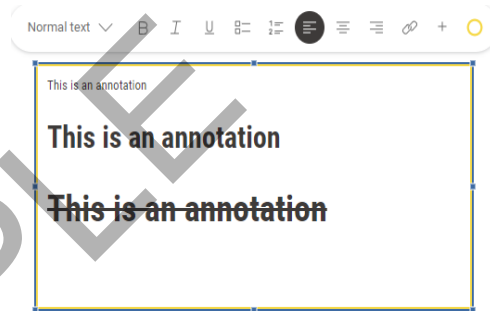



*Figure 1.20. The Annotation Editor.*



## 1.13. Download the KNIME Extensions

KNIME Analytics Platform is an open-source product. As every open-source product, it benefits from the feedback and the functionalities that the community develops. A number of extensions are available for KNIME Analytics Platform. If you have downloaded and installed KNIME Analytics Platform including all its free extensions, you will see the corresponding categories in the Node Repository panel, such as KNIME Labs, Text Processing, R Integration, and many others. However, if at installation time, you have chosen to install the bare KNIME Analytics Platform without the free extensions, you might need to install them separately at some point on a running instance.

## Installing KNIME Extensions

To install a new KNIME extension from within KNIME Analytics Platform, there are three options.

1. From the Top right options, select Menu → "Install KNIME Extensions", select the desired extension, click the "Next" button and follow the wizard instructions.
2. IF you want install extension from the classic UI, From the Top Menu, select "Help" → "Install New Software". In the "Available Software" window, in the "Work with" textbox, select the URL with the KNIME update site (usually named "KNIME Analytics Platform 5.x Update Site" - `http://update.knime.com/analytics-platform/4.x`). Then select the extension, click the "Next" button and follow the wizard instructions.
3. Search the KNIME Community Hub, on a web browser or from the KNIME Community Hub panel. When the desired extension is found, drag and drop the extension icon from the browser to the Workflow Editor.

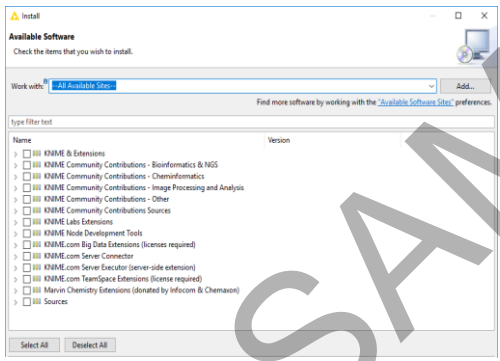



*Figure 1.21. The "Available Software" window.*

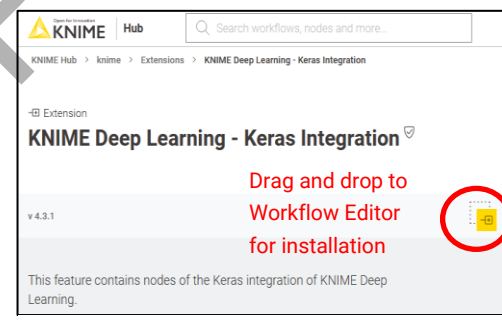



*Figure 1.22. An extension from the KNIME Community Hub.*

Once the selected KNIME extension(s) has/have been installed and KNIME has been restarted, you should see the new category, corresponding to the installed extension, in the "Node Repository".

In the "Available Software" window you can find some extension groups: KNIME & Extensions, KNIME Labs Extensions, KNIME Node Development Tools, Sources, and more. "KNIME & Extensions" contains all extensions provided for the current release; "KNIME Labs Extensions" contains a number of extensions ready to use, but not yet of x.1 release quality; "KNIME Node Development Tools" contains packages with some useful tools for Java programmers to develop nodes; "Sources" contains the KNIME source code. Specific packages donated by third parties or community entities might also be available in the list of extensions. These are usually

grouped under “Community” categories. My advice is to install all extensions, even the cheminformatics ones. Many of them contain several useful nodes not necessarily restricted to a particular domain.

## 1.14. Data and Workflows for this Book

This book builds a few examples and provides the solutions to the exercises. The workflows are accessible via the KNIME Community Hub and are stored in the KNIME Press space – look for the respective book and KNIME version. To download material from the KNIME Community Hub, you need to be logged in with your KNIME account (see how to create a KNIME account). After entering the KNIME Community Hub, in order to download the workflows, just click on the cloud icon. Download the whole folder onto your machine from the KNIME Beginner’s Luck space, which will result in a *.knar* file. Then double click it OR import it into the KNIME Explorer via Select “File” → “Import KNIME Workflow…”.

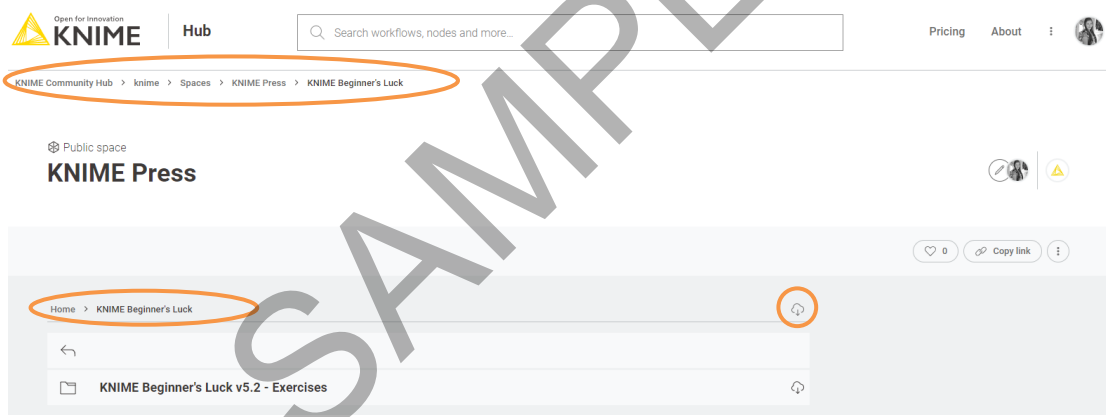



*Figure 1.23. Beginners Luck space on the KNIME Community Hub.*

At the end of the import operation, in the Space Explorer panel you should find a *KNIME Beginner’s Luck v5.2 - Exercises* folder containing Chapter2, Chapter3, Chapter4, Chapter5, Chapter6, and Chapter7 subfolders, each one with workflows and exercises to be implemented in the next chapters. You should also find a KBLdata folder containing the required data.

The data used for the exercises and for the demonstrative workflows of this book were either generated by the author or downloaded from the UCI Machine Learning Repository, a public data repository (http://archive.ics.uci.edu/ml/datasets). If the data set belongs to the UCI Repository, a full link is provided here for download. Data generated by the author, that is not public data, are located in the KBLdata folder.

Data from the UCI Machine Learning Repository:

- *Adult.data:* http://archive.ics.uci.edu/ml/datasets/Adult
- *Iris data:* http://archive.ics.uci.edu/ml/datasets/Iris
- *Yellow-small.data (Balloons):* http://archive.ics.uci.edu/ml/datasets/Balloons
- *Wine data:* http://archive.ics.uci.edu/ml/datasets/Wine

# 1.15. Exercises

## Exercise 1

Create your own workspace and name it “book_workspace”. You will use this workspace for the next workflows and exercises.

### Solution to Exercise 1

- Launch KNIME
- In Workspace Launcher window, click “Browse”
- Select the path for your new workspace
- Click “Launch”

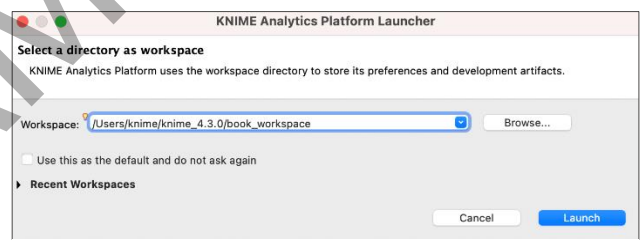



*Figure 1.24. Exercise 1: Create workspace “book_workspace”.*

To keep this as your default workspace, enable the option on the lower left corner.

## Exercise 2

Install the following extensions:

- KNIME Database
- KNIME Javascript Views
- KNIME Report Designer

## Solution to Exercise 2

From the Top right corner options, select Menu → “Install Extensions”. Search and select the required Extensions. Click “Next” and follow the instructions.

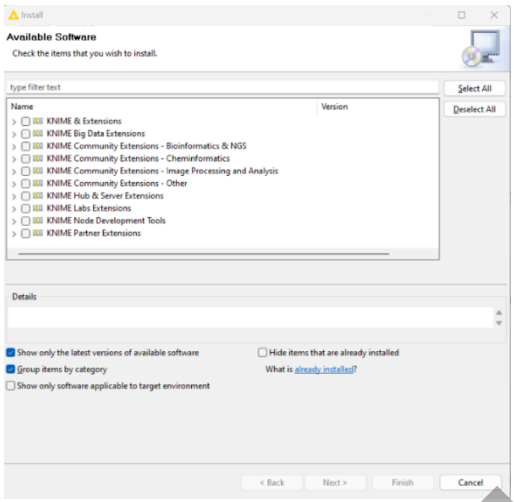



*Figure 1.25. Exercise 2: List of KNIME Extensions.*

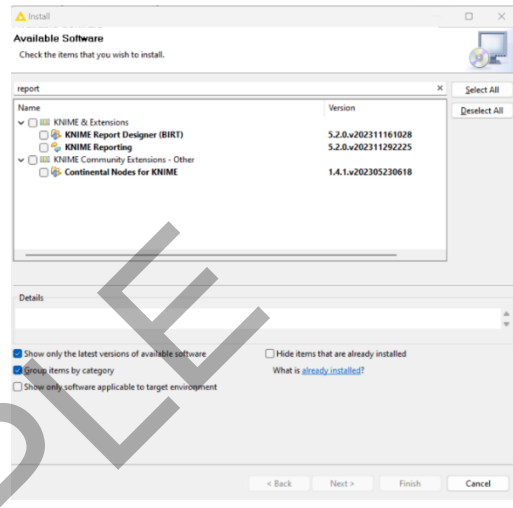



*Figure 1.26. Exercise 2: Reporting Extension.*

# Exercise 3

Search all “Row Filter” nodes in the Node Repository. From the “Node Description” panel, can you explain what the difference is between a “Row Filter”, a “Reference Row Filter”, and a “Nominal Value Row Filter”? Show the node effects by using the following data tables:

**Original Table**

| Position | Name | Team |
|---|---|---|
| 1 | The Black Rose | 4 |
| 2 | Cynthia | 4 |
| 3 | Tinkerbell | 4 |
| 4 | Mother | 4 |
| 5 | Augusta | 3 |
| 6 | The Seven Seas | 3 |

**Reference Table**

| Ranking | Scores |
|---|---|
| 1 | 22 |
| 3 | 14 |
| 4 | 10 |

## Solution to Exercise 3

### Row Filter

The node allows for row filtering according to certain criteria. It can include or exclude certain ranges (by row number), rows with a certain row ID, and rows with a certain value in a selectable column (attribute). In the example below, we used the following filter criterion: `team > 3`

**Original Table**

| Position | Name | Team |
|---|---|---|
| 1 | The Black Rose | 4 |
| 2 | Cynthia | 4 |
| 3 | Tinkerbell | 4 |
| 4 | Mother | 4 |
| 5 | Augusta | 3 |
| 6 | The Seven Seas | 3 |

**Filtered Table**

| Position | Name | Team |
|---|---|---|
| 1 | The Black Rose | 4 |
| 2 | Cynthia | 4 |
| 3 | Tinkerbell | 4 |
| 4 | Mother | 4 |



### Reference Row Filter

This node has two input tables. The first input table, connected to the bottom port, is taken as the reference table; the second input table, connected to the top port, is the table to be filtered. You have to choose the reference column in the reference table and the filtering column in the second table. All rows with a value in the filtering column that also exists in the reference column are kept, if the option “include” is selected; they are removed if the option “exclude” is selected.

**Reference Table**

| Ranking | Scores |
|---|---|
| 1 | 22 |
| 3 | 14 |
| 4 | 10 |

**Filtering Table**

| Position | Name | Team |
|---|---|---|
| 1 | The Black Rose | 4 |
| 2 | Cynthia | 4 |
| 3 | Tinkerbell | 4 |
| 4 | Mother | 4 |
| 5 | Augusta | 3 |
| 6 | The Seven Seas | 3 |

**Resulting Table**

| Position | Name | Team |
|---|---|---|
| 1 | The Black Rose | 4 |
| 3 | Tinkerbell | 4 |
| 4 | Mother | 4 |

In the example above, we use “Ranking” as the reference column in the reference table and “Position” as the filtering column in the filtering table. We have chosen to include the common rows.

**Nominal Value Row Filter**

Filters the rows based on the selected value of a nominal attribute. A nominal column and one or more nominal values of this attribute can be selected as the filter criterion. Rows that have these nominal values in the selected column are included in the output data. Basically, it is a Row Filter applied to a column with nominal values. Nominal columns are string columns and nominal values are the values in it.

In the example below, we use “name” as the nominal column and `name = Cynthia` as the filtering criterion.

**Original Table**

| Position | Name | Team |
|---|---|---|
| 1 | The Black Rose | 4 |
| 2 | Cynthia | 4 |
| 3 | Tinkerbell | 4 |
| 4 | Mother | 4 |
| 5 | Augusta | 3 |
| 6 | The Seven Seas | 3 |

**Filtered Table**

| Position | Name | Team |
|---|---|---|
| 2 | Cynthia | 4 |