

Online-Zertifikatslehrgang

Data Analyst IHK

Die neue Generation digitaler
IHK-Weiterbildungen

INHALT

**VORSTELLUNG DER
LERNZIELE**

TOOLS



**Data
Analyst**_(IHK)

Der Kurs

Ziele und Inhalte

ZIELSETZUNG DES LEHRGANGS

Was können die Teilnehmer nach dem Lehrgang?

- ✓ Teilnehmer, die diesen Kurs erfolgreich absolvieren, bauen sich sowohl ein **fundamentales Wissen** als auch **praxisorientierte Umsetzungsfähigkeiten** in den Bereichen der Datenverarbeitung und Datenanalyse auf.
- ✓ Sie verstehen die **Funktionsweisen von Datenprozessen**, können sie planen und effizient und agil in die Realität umsetzen.
- ✓ Für die Durchführung von Datenanalysen **beherrschen sie intuitive, visuelle Tools** und wissen sie in ihrer betrieblichen Tätigkeit effektiv einzusetzen.
- ✓ Sie können auch **komplexe Daten und Sachverhalte anschaulich und interaktiv visualisieren** und somit ein Maximum an Information zur Verfügung stellen.

MODUL 1: GRUNDLAGEN DER DATA ANALYTICS – DER ETL- PROZESS

Übergreifendes Ziel in Modul 1:

Die Teilnehmer erhalten einen Einstieg in die Grundlagen der Data Analytics sowie Methodenwissen, wie Tools und Plattformen zu bedienen sind.

Thematische Inhalte des Moduls:

- Aufgaben und Funktionen der Data Analytics kennen
- Visuelle Analytics-Werkzeuge verstehen und sicher anwenden
- Den ETL-Prozess verstehen und anwenden
- Mit explorativer Datenanalyse Daten verstehen und visualisieren
- Datenprozesse effizient organisieren und verständlich dokumentieren

Modul 1 – GRUNDLAGEN DER DATA ANALYTICS – DER ETL-PROZESS

1

- **Data Analytics – Definition und Inhalt**
 - Was ist Data Analytics und womit beschäftigt es sich
- **Daten**
 - Datenformen und -quellen
- **Der Datenprozess**
 - Strukturiert von der Datenquelle zur Anwendung
- **Datenimport**
 - Systematisches Extrahieren von Daten
- **Datenqualität**
 - Merkmale der Datenqualität und Methoden zu ihrer Verbesserung

2

- **Die explorative Datenanalyse**
 - Daten zusammenfassen und verstehen
 - Ausreißer und Fehlende Werte entdecken und handhaben
- **Daten bearbeiten und transformieren**
 - Typenkonvertierung, Wertetransformation
 - Tabellentransformation
 - Aggregationen

3

- **Datensicherheit und Datenschutz**
 - Inhalte, Abgrenzungen, wichtige Elemente
- **Datenexport**
 - Daten sichern und Anwendungen zur Verfügung stellen
- **Dokumentation und Organisation von Workflows**
 - Methoden zur Dokumentation und Workflowstrukturierung

MODUL 2: VISUELLE ANALYSE UND REPORTING – EINSTIEG IN BI-TOOLS

Übergreifendes Ziel in Modul 2:

Die Teilnehmer erlangen ein grundlegendes Verständnis zu Funktionen und Aufgaben von Business Intelligence Tools. Übungen und praxisnahe Beispiele ermöglichen den Teilnehmern, die Vorteile von visuellen Analysen und Reportings effizient im Beruf anzuwenden.

Thematische Inhalte des Moduls:

- Aufgaben und Funktionen von visueller Analyse und Reporting kennen
- BI-Tools verstehen und sicher anwenden
- Inhalte und Daten effizient und verständlich visualisieren

Modul 2 – VISUELLE ANALYSE UND REPORTING - BI TOOLS

1

- **Was ist Business Intelligence?**
 - Einsatz und Funktionen
- **Das Tool im Kurs**
 - Kurze Einweisung in die Handhabung
- **Datenimport**
- **Datenverbindungen**
- **Berechnete Spalten und Measures**
- **Diagramme und Ansichten**
 - Daten bildlich darstellen

2

- **Filter**
 - Inhalte fokussieren und Aufmerksamkeit lenken
- **Dashboards**
 - Wichtige Inhalte auf eine Seite konzentrieren
 - Dashboard-Typen und ihre Anwendung
 - Dashboards systematisch mit dem Dashboard Cycle entwickeln

MODUL 3:

DATA ANALYTICS FÜR
FORTGESCHRITTENE – DATENBANKEN,
MACHINE LEARNING, WORKFLOW
CONTROL

Übergreifendes Ziel in Modul 3:

Die Teilnehmer erhalten ein tieferes Verständnis, Methodenwissen und einen sicheren Umgang mit der Datenanalytik. Anhand anschaulicher Datensätze lernen sie, Datenmodelle zu erstellen und zu optimieren.

Thematische Inhalte des Moduls:

- Sicherer Umgang und Arbeiten mit Datenbanken
- Verschiedene Typen des Maschinellen Lernens und deren Anwendungsgebiete verstehen
- Kenntnisse zu Datenmodellierung und Maschinellern Lernen umsetzen und für eigene Datenmodelle anwenden
- Methoden für Strukturierung und Kontrolle von Workflows anwenden

Modul 3 - DATA ANALYTICS FÜR FORTGESCHRITTENE – DATENBANKEN, MACHINE LEARNING, WORKFLOW CONTROL

1

- **Workflowkontrolle**
 - Flow Variablen
 - Benutzerinterfaces
 - Schalter
 - Schleifen

2

- **Datenbanken**
- **Fehlermanagement und Streaming**

3

- **ML und KI Teil 1**
 - Übersicht
 - Überwachtes Lernen
 - Unüberwachtes Lernen

4

- **ML und KI Teil 2**
 - Neuronale Netze
 - Modelle bewerten
 - Modelle optimieren

MODUL 4:

(EIGENE) DATENPROJEKTE – BEWERTEN, PLANEN UND UMSETZEN

Übergreifendes Ziel in Modul 4:

Die Teilnehmer lernen Datenprojekte zu bewerten und zu begründen sowie anhand agiler Methoden in einzelnen Teilschritten und Versionen zu planen und durchzuführen.

Thematische Inhalte des Moduls:

- Verstehen, wie man Datenprojekte sicher plant und richtig argumentiert
- Projektparameter und -ziele anschaulich formulieren und visualisieren
- Agile Methoden anwenden, um flexible Datenprojekte effizient zu realisieren

Modul 4 – Datenprojekte – bewerten, planen und umsetzen

1

➤ Grundlagen des Agilen Arbeitens

➤ Einleitung und Ist-Analyse:

- Beschreibung der Ausgangssituation und des Hintergrundes der Daten
- Erläuterung der Zusammensetzung der Daten
- Bewertung der Datenqualität

➤ Backlog und Projektziele

- Anforderungen: User Storys und Potenziale
- Bewertung von Aufwand und Ertrag der Elemente im Backlog
- Priorisierung der Backlogelemente und Auswahl der erreichbaren Projektziele

2

➤ Hypothesen:

- Überprüfung der Machbarkeit der Funktionen für die User Storys und Potenziale
- Zuordnung der Hypothesen zu User Storys und Potenzialen

➤ Metriken:

- Wie lassen sich Fortschritt und Erfolg des Projektes und die Qualität der Analyse bzw. des Dashboards im Einsatz messen?
- Definitions of done: Akzeptanzkriterien für das Erreichen der ausgewählten Ziele

3

➤ Machine Learning Canvas:

- Funktionale und technische Projektübersicht des ersten Sprints

➤ Planung des ersten Sprints:

- Formulierung der einzelnen Arbeitsschritte und Aufgaben sowie des dafür benötigten Aufwandes

MODUL 5:

PRAKTISCH UMSETZEN –
DATENPROJEKTE PLANEN,
MODELLE ERSTELLEN UND
VORHERSAGEN TREFFEN

Übergreifendes Ziel in Modul 5:

Die Teilnehmer sind in der Lage, bei der Bearbeitung von betrieblichen Problemstellungen eigene Lösungsstrategien in der Datenanalyse zu entwickeln und umzusetzen. Sie nutzen Data Analytics für ihre eigene Weiterentwicklung.

Thematische Inhalte des Moduls:

- Wissen über Data Analytics im beruflichen Kontext anwenden
- Anforderungen und Ziele für sich ständig ändernde Aufgabenstellungen darstellen und formulieren
- Verbesserungspotenzial durch Datenanalytik erkennen und entsprechende Entwicklungen anstoßen

PROJEKTARBEIT UND ABSCHLUSSPRÄSENTATION

- ✓ **Die Teilnehmer führen eine lehrgangsbegleitende Projektarbeit durch.**
Diese Arbeit soll das Verständnis für die Anwendung der Themen aufzeigen. In der Projektarbeit soll daher ein Thema aus dem Arbeitsalltag systematisch aufbereitet werden. Im Vordergrund steht der Praxistransfer in die betriebliche Realität. Im Zuge der Implementierung eines (kleinen) Datenprojektes sollen sowohl der Nutzen als auch die persönlichen Lernerfahrungen der Teilnehmer thematisiert werden.
- ✓ **Die Projektarbeit wird in Kleingruppen durchgeführt.** Bei Gruppenarbeiten ist der Verantwortungsbereich der Teilnehmer genau zu definieren, so dass der Beitrag jedes Einzelnen klar erkennbar ist.
- ✓ **Die mündliche Präsentation der Projektarbeit wird 30 Minuten dauern**
(Projektplanung und Ausarbeitung und Demonstration des Workflows und Ergebnisvorführung) und durch Fragen der Testkommission abschließend geprüft werden.

Inhalt der Projektarbeiten (pro Gruppe)

Projektbeschreibung

1. Einleitung und Ist-Analyse:

- Beschreibung der Ausgangssituation und des Hintergrundes der Daten (Unternehmen, Branche, Prozesse, etc.).
- Erläuterung der Zusammensetzung der Daten
- Bewertung der Datenqualität.

2. Backlog und Projektziele:

- Sammlung von User Storys und Potenzialen.
- Bewertung und Priorisierung nach Aufwand und Ertrag (Low Hanging Fruits)
- Auswahl der erreichbaren Projektziele für den ersten Sprint.

3. Projektausarbeitung:

- Formulierung von Hypothesen, die die Machbarkeit der Funktionen der User Storys und Potenziale untermauern.
- Zuordnung der Hypothesen zu User Storys und Potenzialen.

4. Projektmetriken:

- Wie lassen sich Fortschritt und Erfolg des Projektes und die Qualität der Analyse bzw. des Dashboards im Einsatz messen?
- Definitions of done für die Mehrwerte der ausgewählten Ziele.

5. Planung des ersten Sprints:

- Formulierung der einzelnen Aufgabenpakete mit ihrem berechneten oder geschätzten Aufwand.

6. Technische Beschreibung des ersten Sprints anhand des Machine Learning Canvas.

7. Ergebnisse der EDA und der Hypothesenüberprüfung.

8. Optional: Bedienungsanleitungen und Ergänzungen zur Dokumentation (Dashboard bzw. Workflow).

→ Abgabe: Dokument „Projektbeschreibung“ (PDF)

Inhalt der Projektarbeiten (pro Gruppe)

Datenanalyse

- Strukturiertes Workflow (Datenimport, Datenbereinigung, Datentransformation, Datenanalyse, Datenexport)
- EDA zu den importierten Daten
- Datenanalyse mit Bewertung, Optimierungs- und Stabilitätstest.
- Einfache (grafische) Darstellung der Ergebnisse

→ Abgabe: Workflow mit Dokumentation und Input-Daten als Archiv (.knar)

Datenvisualisierung

- Bericht oder Dashboard mit Visualisierung der Daten
- Interaktive Elemente zur flexiblen Erkundung der Informationen
- Darstellung des Zusammenhangs der Informationen (Welche Geschichte soll erzählt werden?)
- Kurze Dokumentation der Datenbearbeitung

→ Abgabe: Interaktiver Bericht (.pbix) mit kurzer Funktionsbeschreibung und Input-Daten

The background of the slide features a blue-toned image of a hand reaching out towards a wireframe globe. The globe is composed of a network of white lines and dots, symbolizing global connectivity and digital technology. The hand is positioned in the lower-left foreground, with fingers slightly curled as if about to touch the globe.

Tool im Kurs: KNIME

Was ist die KNIME Analytics Plattform?

Ein open source Data Science Tool mit graphischer Oberfläche für:

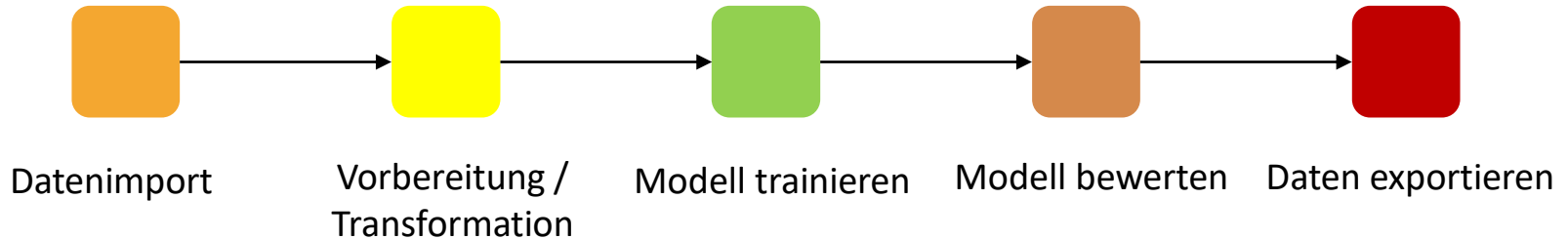
- Extraktion
- Manipulation
- Analyse
- Darstellung
- Reporting von Daten

Mit vielen Community-basierten Extensions in den Bereichen:

- Text Mining
- Big Data
- Cheminformatics
- Deep Learning
- Integration und Schnittstellen zu anderen Systemen wie R, Python, Tableau etc.

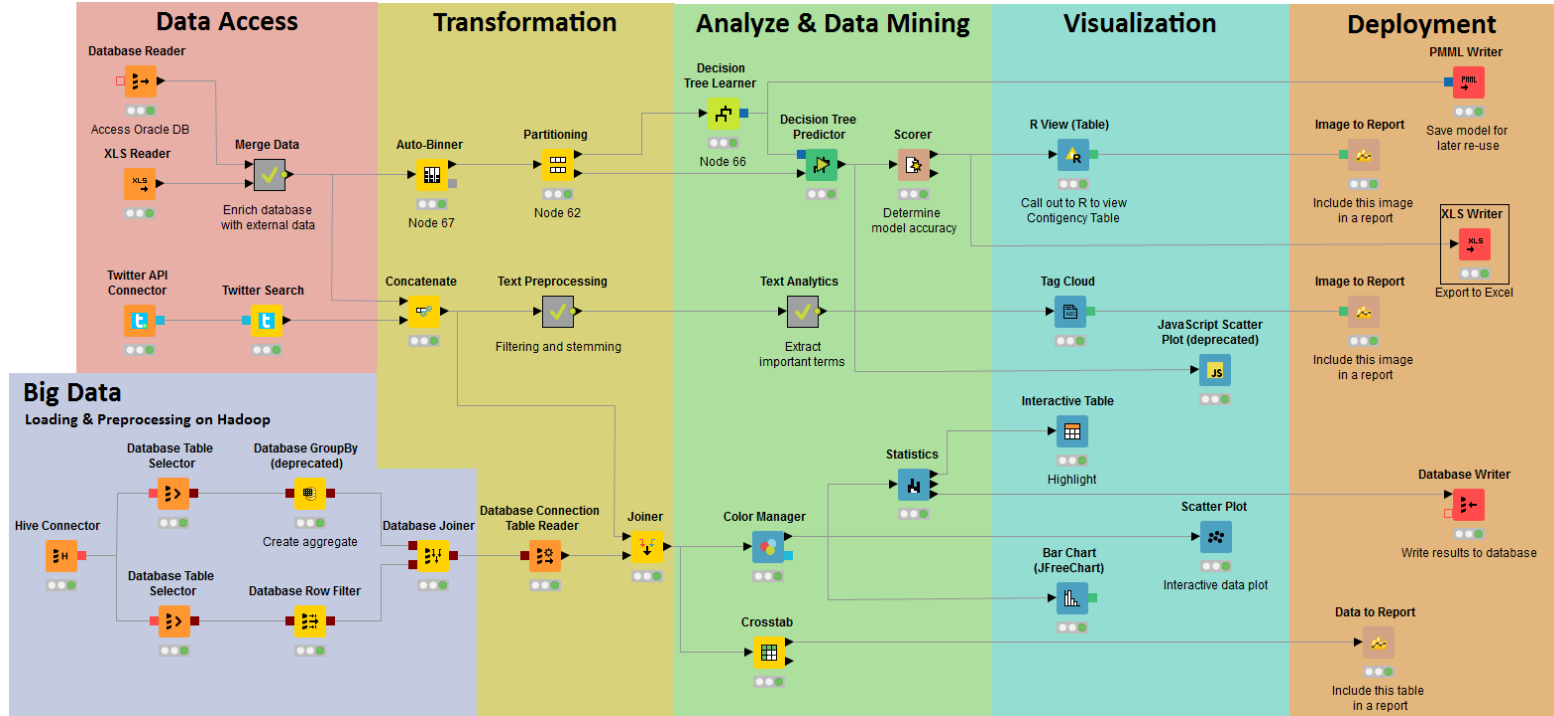
Warum verwenden wir KNIME?

- Durch die Anordnung von Knoten (Nodes) können wir alle Schritte einer Analyse abbilden und umsetzen, ohne eine Programmiersprache zu verwenden.

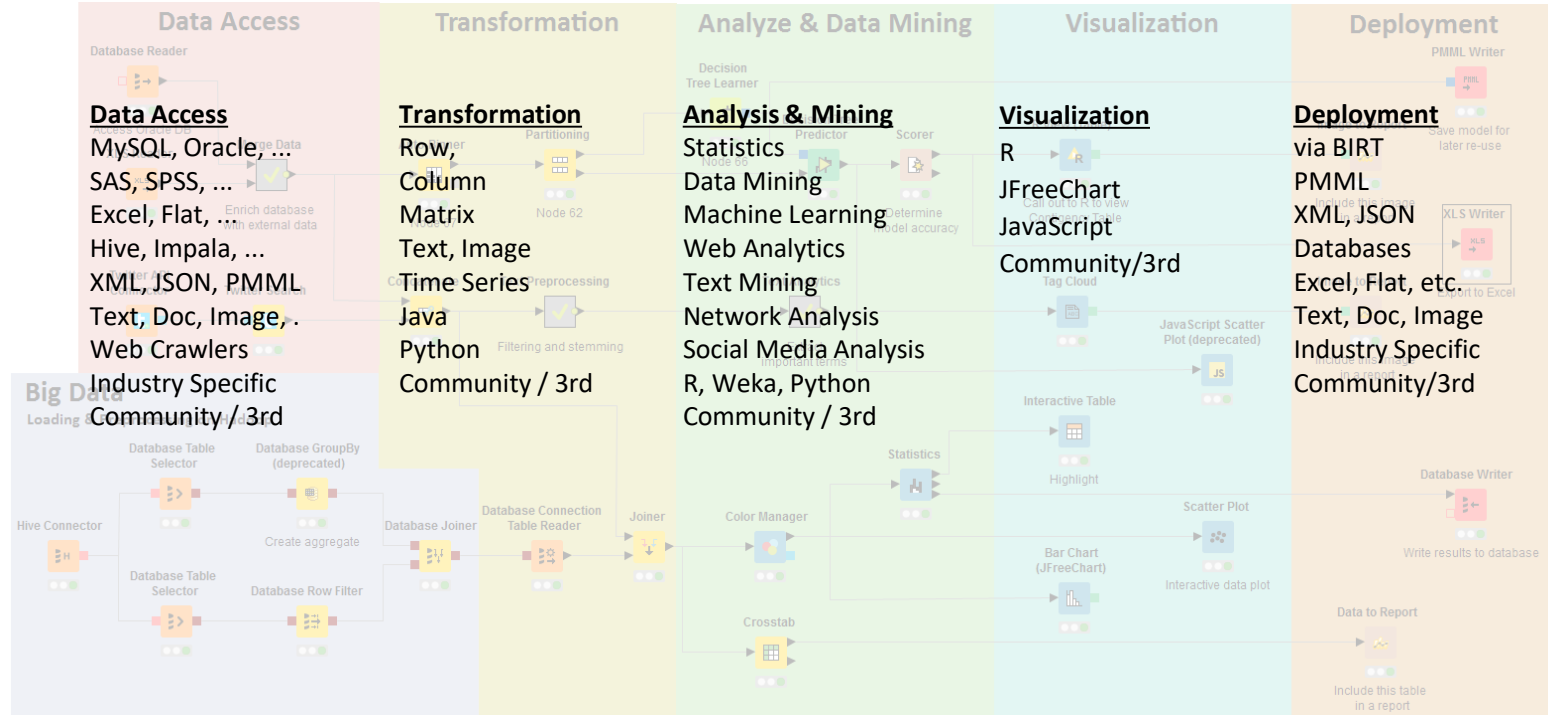


- Jeder Baustein hat eine definierte Aufgabe.
- Bsp: Eine Spalte in zwei Spalten aufteilen

Es gibt über 2000 farblich gekennzeichnete Knoten

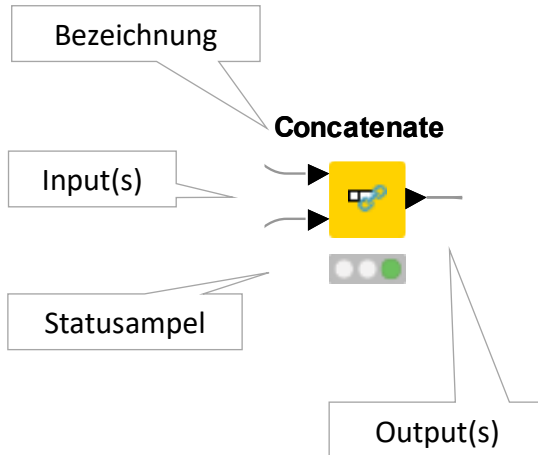


Es gibt über 2000 farblich gekennzeichnete Knoten

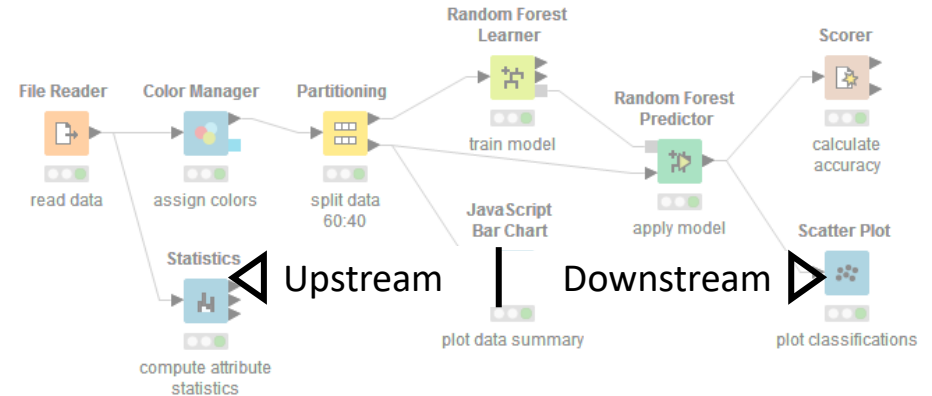


Knoten und Workflows

- **Aufbau:**



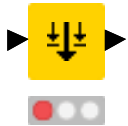
- Knoten werden zu **Workflows** kombiniert.
- Linien zeigen den **Fluss der Daten** von Knoten zu Knoten.



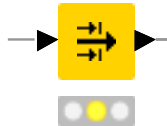
Statusampel

- rot:** unkonfiguriert, Knoten muss konfiguriert werden
► Konfigurations-menü, rechtsklick, oder F6
- gelb:** konfiguriert, fertig für die Ausführung
- grün:** ausgeführt, Ergebnis liegt am Output vor
- gelbes Dreieck:** Konfiguration oder Ausführung mit Warnung
- roter Kreis:** Fehler, Knoten konnte nicht ausgeführt werden

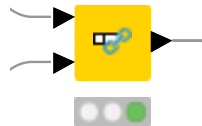
Column Filter



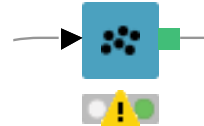
Row Filter



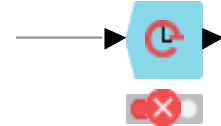
Concatenate



Scatter Plot
(JFreeChart)



TimeDelay Loop End



Installation



Installation:

- Die kostenlose KNIME Analytics Plattform gibt es unter: <https://www.knime.com/downloads/download-knime>
- Download der aktuellen Version (Installer - falls Admin; zip-Archive – wenn kein Admin)
- Installieren Sie das Programm / Entpacken Sie das zip-File

Erster Start:

- Man muss einen KNIME Workfolder und ein Workspace erstellen.
Dies sind die Ordner in denen später die KNIME Workflows gespeichert werden

KNIME Oberfläche („workbench“)

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a welcome message and a workflow diagram. The left sidebar contains the KNIME Explorer, Workflow Coach, and Node Repository. The right sidebar shows the Node Description for a 'WrappedNode Output'.

KNIME Explorer

- EXAMPLES (knime-guest@http://public-server)
- LOCAL (Local Workspace)
 - Analysis_SSL
 - BoschSI
 - ETK_Prognose_Knoll
 - Example Workflows
 - GuK_Audit
 - KNIME Server
 - References
 - Showcase_2017
 - Data Showcase
 - Showcase_2017
 - TrainingsAndCourses
 - Wahlfrei Dieral

Workflow Coach

Recommended Nodes

Node Repository

- IO
- Manipulation
- Views
- Analytics
- Database
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Community Nodes
- KNIME Labs
- Workflow Control

Welcome to KNIME Analytics Platform!

New to KNIME? Looking for resources to get started?

- Register for emails with introductory tips [here](#).
- Explore our [Checklist Guide](#).
- Check out [Tutorials](#) to get started with KNIME Analytics Platform.
- Find more hints and how-tos in the [Learning Hub](#).
- And register for our release and event emails [right here](#).

This page will be displayed upon startup but you can customize the content using the checkboxes at the bottom.

Updates for the following components are available:

- DPHAT/UCI Open Modeling Extensions
- Pullman for KNIME

Click [here](#) in order to install updates.

Where to go from here

- Create new workflow
- Learning Hub
- Browse example workflows
- Get additional nodes
- Go to my workflows
- Mount KNIME Cloud Server

Most recently used workflows

- ModelSelection_WebPortal_Part1
- ModelSelection_WebPortal_Part1
- ModelSelection_BasicWorkflow
- DataCleaning_WebPortal_v2.0
- KNIME_jmpact2
- Silly ETL_v2.0

Tips & Tricks

Specialist Nodes

Did you know there are a whole variety of specialist nodes available from KNIME Labs and the Community around Scripting, Image Processing, Text Processing, Internet Mining, Network Mining, Cell Biology and Genetics, and Chemistry. To access them, go to Help Menu and choose Install New Software.

- ☒ Show intro text at next start
- ☒ Show update notifications at next start
- ☒ Show links and most recently used workflows at next start

Node Description: WrappedNode Output

WrappedNode output handle. All objects passed to this node will be forwarded downstream of the WrappedNode.

Ports

This node is contained in *KNIME Core* provided by *KNIME AG, Zurich, Switzerland*.

KNIME Console

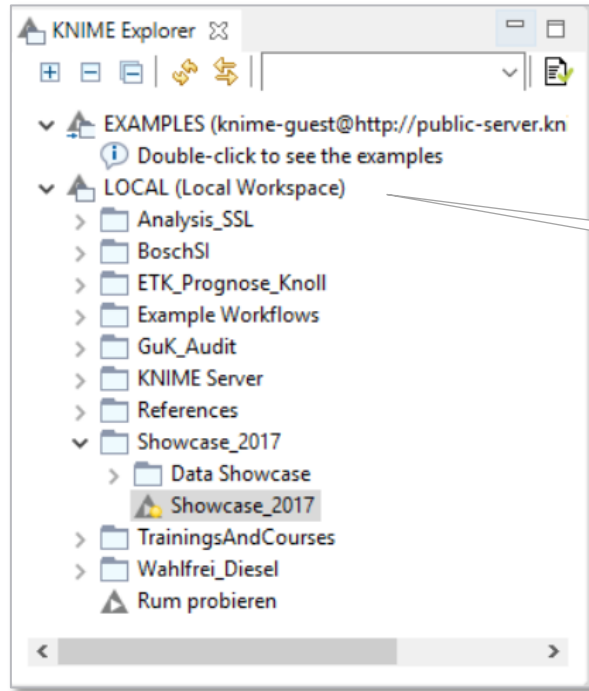
```
*** Welcome to KNIME Analytics Platform v3.6.0.v201807100937 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: D:\Data Science\KNIME\metadata\knime\knime.log
```

KNIME Oberfläche („workbench“)

The screenshot displays the KNIME Analytics Platform interface, which is a data science workbench. The interface is divided into several panes and sections, each highlighted with a blue label:

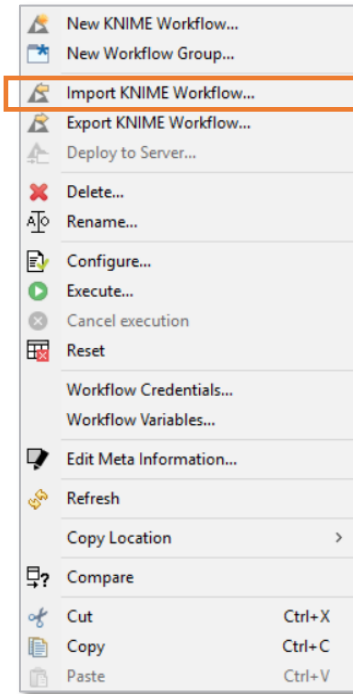
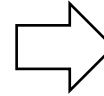
- KNIME Explorer**: Located in the top-left pane, it shows a tree view of the project structure, including folders like 'EXAMPLES' and 'Data Showcase'.
- Workflow Coach**: A central pane that provides guidance and recommendations for building workflows. It includes sections like 'Welcome to KNIME Analytics Platform!', 'Where to go', and 'Tips & Tricks'.
- Node Repository**: Located in the bottom-left pane, it provides a searchable list of available nodes categorized by function (e.g., IO, Manipulation, Views, Analytics, Database).
- Workflow Übersicht**: A central pane that displays the current workflow canvas, showing nodes and their connections.
- Node Beschreibung**: A pane on the right side that provides detailed information about the selected node, including its purpose and usage.
- Meldungen (Console)**: A pane at the bottom right that displays system messages and logs.

KNIME Explorer

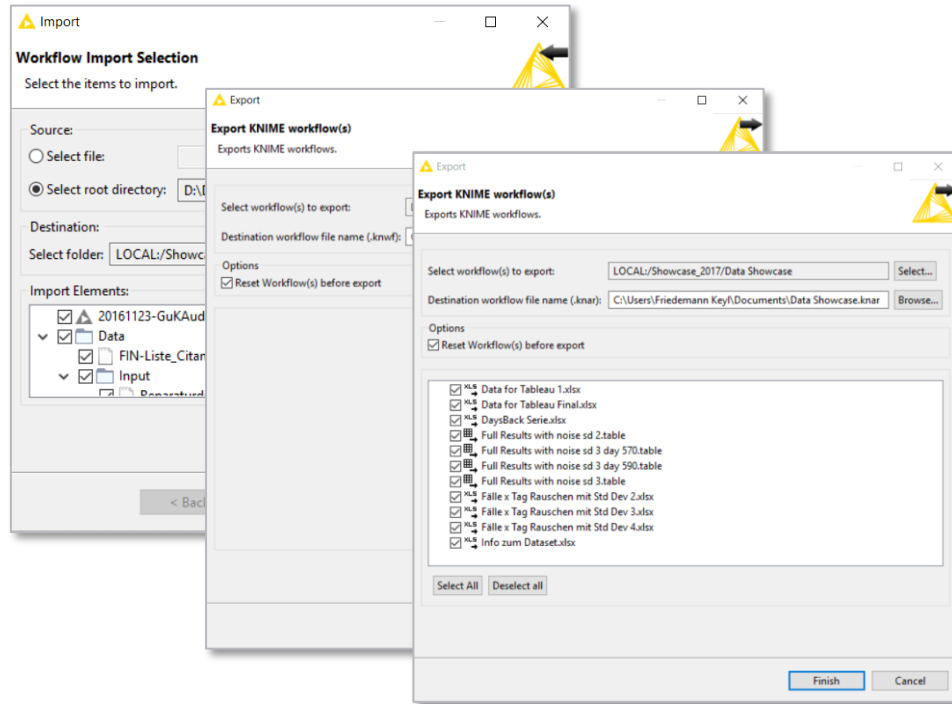


Beispiele auf dem
Public Server

Lokaler Ordner



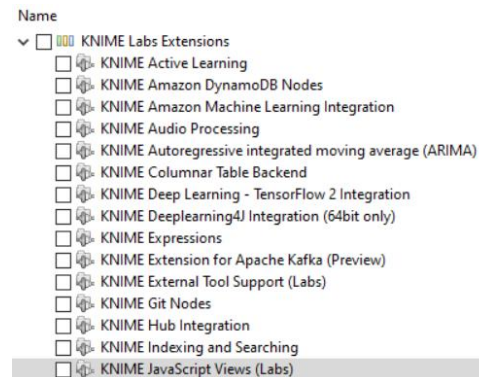
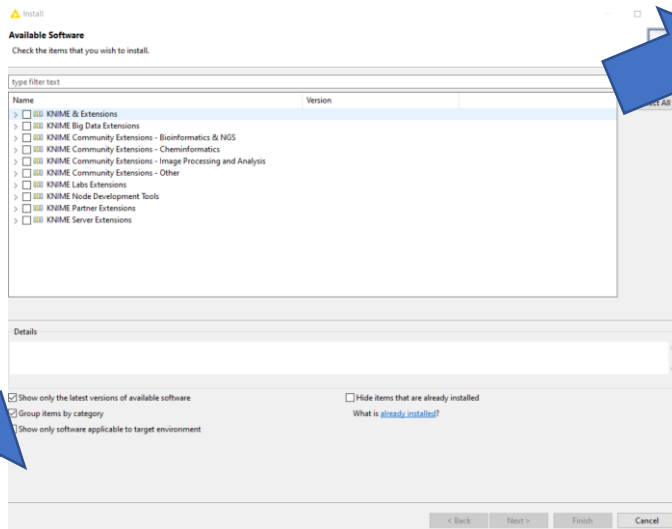
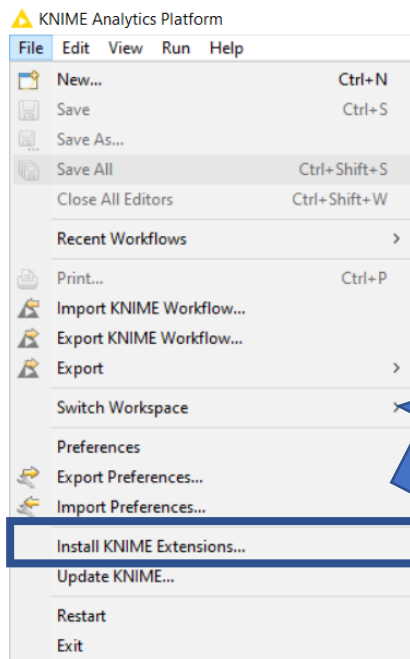
Neue Workflows, Import und Export



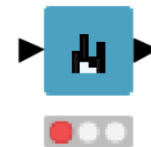
- KNIME Dateityp:
- ***.knwf**
für KNIME Workflows
- ***.knar**
für KNIME Archive
(enthält Workflows und Daten)

Extensions

KNIME lässt sich modular erweitern:



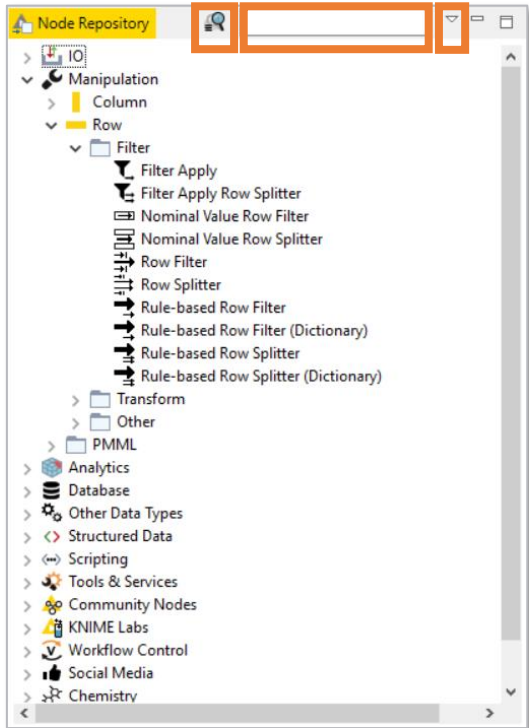
Data Explorer



Liste der KNIME Extensions, die im Laufe des Kurses benötigt werden:

- KNIME Data Generation
- KNIME JavaScript Views (Labs)
- KNIME REST Client Extension
- KNIME JSON Processing
- KNIME Quick Forms
- KNIME Ensemble Learning Wrappers
- KNIME Distance Matrix
- KNIME Optimization Extension
- KNIME Plotly

Node Repository



- Enthält alle verfügbaren Knoten
- Geordnet nach Kategorien
- Suche nach Nodes über „crisp“ oder „fuzzy“ Search
- Einfügen der Knoten in den Workflow im aktuell aktiven Workflow Editor über:
 - drag & drop oder
 - Doppelklick auf den gewünschten Knoten

Node Beschreibung

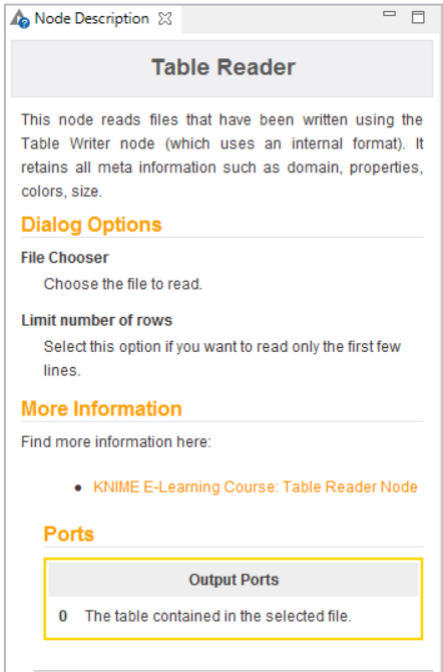


Table Reader

This node reads files that have been written using the Table Writer node (which uses an internal format). It retains all meta information such as domain, properties, colors, size.

Dialog Options

File Chooser
Choose the file to read.

Limit number of rows
Select this option if you want to read only the first few lines.

More Information
Find more information here:

- [KNIME E-Learning Course: Table Reader Node](#)

Ports

Output Ports	
0	The table contained in the selected file.

Bezeichnung des Knotens

Beschreibung des Knotens

Beschreibung der Konfigurationsoptionen

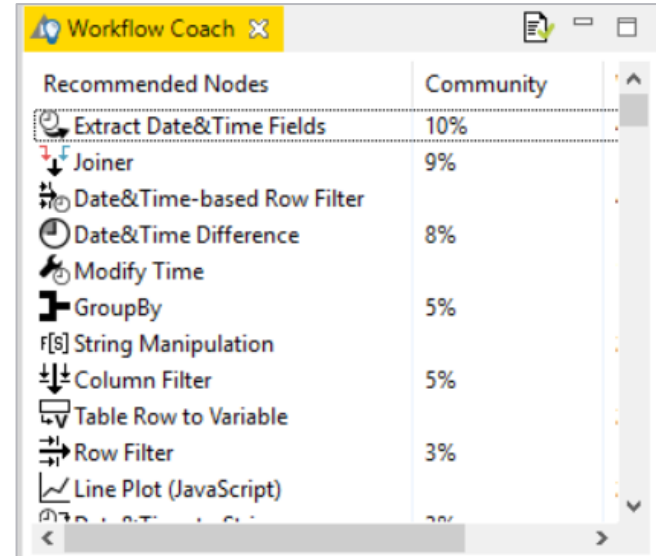
Weitere Informationsquellen

Erklärung der Input und Output Ports des Knotens

- Anklicken eines Knotens zeigt die Beschreibung im Bereich „Node Description“ an.
- Info zu allen Knoten auch im Node Guide, dort mit Beispiel Workflows:
<https://www.knime.com/nodeguide>
- Und im Node Pit:
<https://nodepit.com/>

Workflow Coach

- Zeigt für jeden Knoten die nachfolgend am häufigsten verwendeten Knoten
- Wie im Repository per drag & drop oder Doppelklick zum Workflow hinzufügen
- Analysiert die Nutzung von Knoten in den Workflows aller User
- Muss extra freigegeben werden
- Auch nur lokale Nutzung möglich...

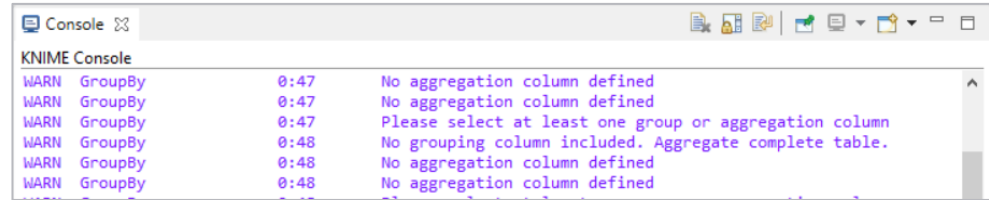


The screenshot shows the 'Workflow Coach' window with a table of recommended nodes. The table has two columns: 'Recommended Nodes' and 'Community'. The nodes are listed in descending order of community usage percentage.

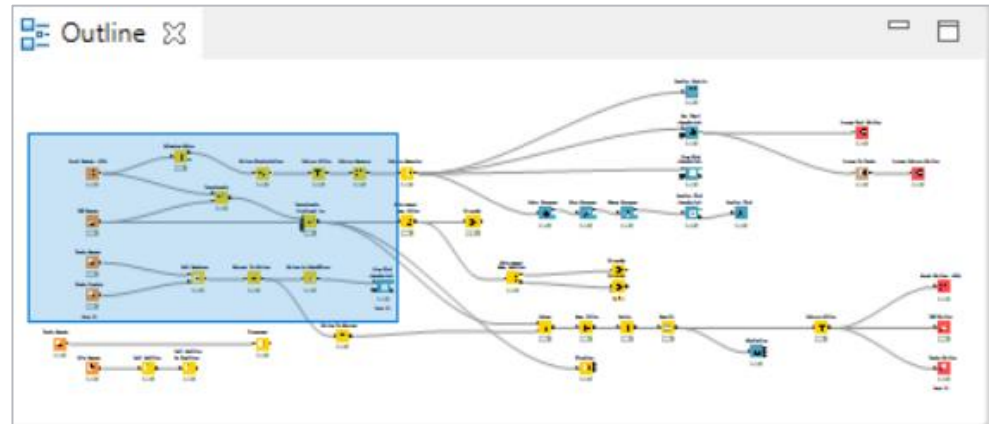
Recommended Nodes	Community
Extract Date&Time Fields	10%
Joiner	9%
Date&Time-based Row Filter	8%
Date&Time Difference	8%
Modify Time	5%
GroupBy	5%
f[s] String Manipulation	5%
Column Filter	5%
Table Row to Variable	3%
Row Filter	3%
Line Plot (JavaScript)	3%

Konsole und Outline

Die Konsole (Console) zeigt Warn- und Fehlermeldungen

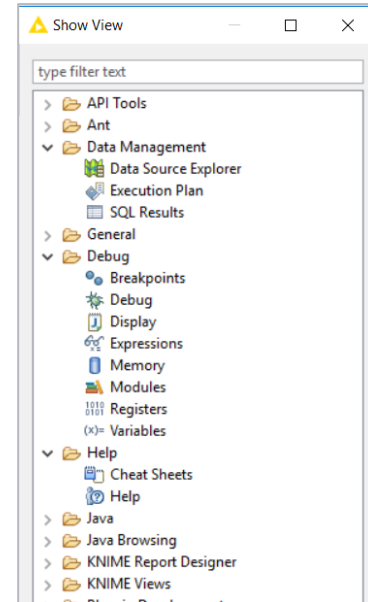
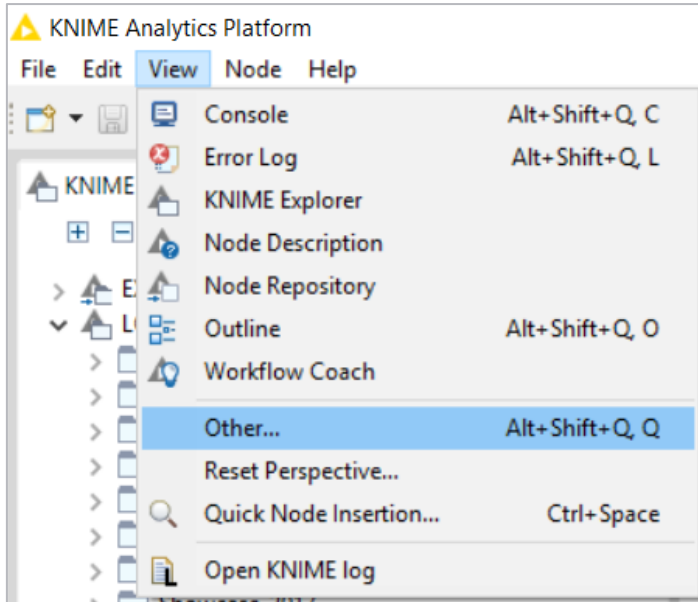


Workflowübersicht (Outline) zeigt auch bei großen Workflows die gesamte Sicht sowie den im Workflow Editor momentan dargestellten Ausschnitt.



Weitere Views sind individuell einstellbar...

- Eine große Anzahl weiterer Ansichten finden sich im Reiter View:



Toolbar

