

Rebecca Schwarz

# GSC-DATEN AUFBEREITEN: Vorbereitung einer SEO-Analyse mit KNIME und Google Sheets

Ohne solide Datengrundlage ist keine erfolgreiche Analyse möglich. Deshalb soll in dieser Ausgabe anhand eines einfachen SEO-Praxisbeispiels veranschaulicht werden, wie die Open-Source-Software KNIME Tasks zur Datenaufbereitung ermöglicht und wiederholbar macht. Konkret wird gezeigt, wie Leistungsdaten mithilfe von Google Sheets in der Google Search Console abgefragt, mithilfe von KNIME die Top-Suchanfragen je Seite ermittelt und klassifiziert werden. Der Prozess endet mit einem automatisierten Export in ein neues Google Sheet zur weiteren individuellen Bearbeitung. Ziel ist es, die vielfältigen Möglichkeiten von KNIME in der Datenvorbereitung zu zeigen, weshalb alle beschriebenen Schritte jederzeit anpassbar und erweiterbar sind.

## Sinnvolle SEO-Analysen benötigen eine solide Datenvorbereitung

Bei der Analyse von Leistungsdaten einer Website führt kein Weg an der Google Search Console (kurz: GSC) vorbei. Die GSC ist die Kommandozentrale der Website und erhebt alle Daten, die mit der organischen Performance der Website zu tun haben. Sollen neue Erkenntnisse über die Leistung vieler URLs der Website gewonnen werden, kommt die Arbeit mit dem GSC-Frontend jedoch schnell an ihre Grenzen.

Folgender Anwendungsfall soll das verdeutlichen: Für eine umfangreiche Analyse soll das Low-Hanging-Fruit-Potenzial von verschiedenen URLs erhoben werden. Dazu werden die fünf meistgeklickten Suchanfragen je URL betrachtet. Zusätzlich soll dann eingeordnet werden, welche dieser Suchanfragen bei einer durchschnittlichen Position zwischen vier und

elf in den Suchergebnissen liegen und zugleich mehr als den Durchschnitt an Impressionen des gesamten Datensatzes erzielen.

**Spoiler:** Ohne händische Klickarbeit im GSC-Frontend und die ständige Ansichtsänderung im Tool zwischen Seite zu Suchanfrage lässt sich dieses Aggregationslevel nicht darstellen. Auch über einen Datenexport ist das nicht so einfach möglich, da auch hier Seiten und die zugehörigen Suchanfragen nicht zusammen aufgeführt werden.

**Aber gute Nachrichten:** Weil diese Ansicht für eine SEO-Datenanalyse immer wieder gefragt ist, gibt es hierfür ein eigenes Chrome-Plug-in. Mithilfe des nachfolgenden Workflows werden die GSC-Daten dann in KNIME aufbereitet und klassifiziert. Das Ergebnis ist eine Tabelle, die URLs mit den zugehörigen Top-Suchanfrage beinhaltet und nach individueller Klassifizierung

Foto: natrot / gettyimages.de

### DIE AUTORIN



Rebecca Schwarz ist SEO-Consultant bei der getraction GmbH. Ihr Schwerpunkt liegt in der Entwicklung von SEO-Strategien und der Durchführung von Content-Audits. Um größere Datenmengen effizient zu verarbeiten und um bei wiederkehrenden SEO-Tasks Zeit zu sparen, nutzt sie die Open-Source-Software KNIME.

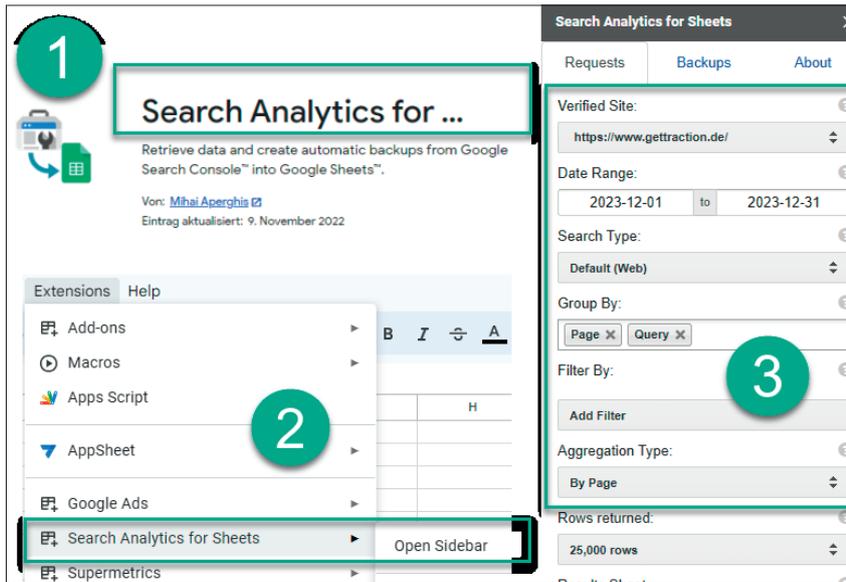


Abb. 1: Konfiguration des Chrome-Plug-ins „Search Analytics for Sheets“

den Kriterien für Low-Hanging-Fruit-Potenzial entspricht. Die Tabelle kann schlussendlich automatisch in Google Sheets exportiert werden und zur individuellen Auswertung weiterverwendet werden.

Das alles wäre ohne KNIME nicht möglich! Denn mithilfe von KNIME werden auch in größeren Datensätzen Potenziale sichtbar und das Beste: Ist der Workflow einmal konfiguriert, lassen sich Daten der gleichen Struktur immer wieder verarbeiten. Der vorgestellte Workflow ist so gestaltet, dass er sich leicht anpassen lässt und damit vielfältig einsetzbar wird. Die Anleitung zeigt nun folgende Schritte:

- » Export des geforderten Aggregationslevels „Suchanfragen je URL“ aus der GSC
  - » Laden der Daten via Google Sheets in KNIME
  - » Aufbereitung der Daten nach Top-Suchanfragen je URL
  - » Klassifikation der Suchanfragen nach Low-Hanging-Fruit-Potenzial
  - » Export der aufbereiteten Daten aus KNIME zurück in Google Sheets
- Benötigt werden für die Ausführung folgende Tools:
- » Zugang zur GSC
  - » Chrome-Plug-in „Search Analytics for Sheets“
  - » KNIME-Umgebung

### Schritt eins

Zum Start werden die GSC-Daten im Aggregationslevel „Suchanfragen je URL“ benötigt. Diese Ansicht kann kostenfrei mithilfe des Browser-Plug-ins „Search Analytics for Sheets“ aus der GSC exportiert werden. Dazu muss das Chrome-Plug-in im Chrome Web Store (Abbildung 1, Ziffer 1) installiert werden. Um das Plug-in zu starten, wird ein leeres Google Sheet geöffnet und die Plug-in-Oberfläche über Extensions → „Search Analytics for Sheets“ → „Open Sidebar“ aufgerufen (Abbil-

dung 1, Ziffer 2). Für die Konfiguration sind nun die entsprechende Domain, der gewünschte Zeitraum der Daten und der Suchtyp auszuwählen. Danach kommt die entscheidende Einstellung, die Wahl des Aggregationslevels. Hierbei werden die Daten zunächst nach „page“ und „query“ gruppiert und dann auf die Aggregation „by page“ gebracht (Abbildung 1, Ziffer 3). Google stellt unter [einfach.st/knime647](https://einfach.st/knime647) eine ausführliche Dokumentation der Datenstruktur der GSC zur Verfügung. Das Ergebnis ist eine Tabelle, die je Seite und Suchanfrage eine eigene Zeile beinhaltet. Rankt eine Seite zu beispielsweise zehn unterschiedlichen Suchanfragen, existieren zur Seite zehn Zeilen, die aggregiert auf die Suchanfrage die organischen Klicks, Impressionen, CTR und die durchschnittliche Position zeigen.

### Schritt zwei

Nun beginnt die Arbeit in der Datensoftware KNIME. In früheren Ausgaben wurde bereits gezeigt, wie Daten mithilfe von READER-Knoten in die Oberfläche geladen werden können. Diesmal werden die Daten direkt aus Google Sheets importiert, denn

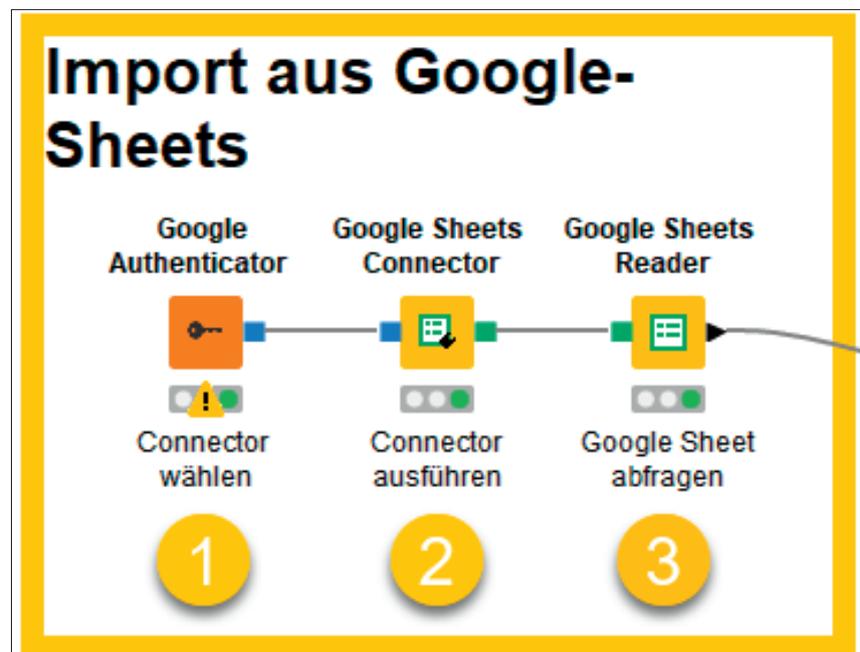


Abb. 2: Knoten zum Import der Daten aus Google Sheets

in KNIME gibt es dafür vorgefertigte Knoten. Zum Datenimport sind drei GOOGLE-SHEETS-Knoten notwendig (Abbildung 2).

Als Erstes wird der GOOGLE-AUTHENTICATOR, in dem die Authentifizierung an Google Sheets läuft, konfiguriert. Dazu muss der entsprechende Chrome-Browser, in dem das Google Sheet erstellt wurde, geöffnet sein. Der Prozess startet mit Doppelklick auf den Knoten und Klick auf „Login“ (Abbildung 3). Es öffnet sich automatisch der Default-Browser, in dem der Zugriff für KNIME erteilt werden kann. Handelt es sich bei diesem Browser nicht um den gewünschten, kann einfach die URL kopiert und im Browser, in dem Zugriff auf das Google Sheet besteht, geöffnet werden.

Danach wird der GOOGLE-SHEETS-CONNECTOR mit dem ersten Knoten verbunden, hier ist keine Konfiguration notwendig. Ein Rechtsklick auf den Knoten und die Auswahl „Execute“ genügen. Verändert sich das Ampelsymbol unterhalb des Knoten von Gelb zu Grün, war die Ausführung erfolgreich und der dritte Knoten kann verbunden werden. Denn zum Einlesen der Daten fehlt nun noch der GOOGLE-SHEETS-READER. In der Konfiguration muss über „SELECT“ das entsprechende Datensheet ausgewählt werden. Außerdem sollte ein Haken bei „Has Column Header“ gesetzt werden, um die Spaltenbenennung des Sheets zu übernehmen. Wichtig: Der Haken bei „Has Row Headers“ darf nicht gesetzt werden.

**Schritt drei**

Nun ist ein Zwischenschritt notwendig, um mit den Daten korrekt weiterarbeiten zu können. Denn KNIME erkennt beim Einlesen der Daten die Werte als Datenformat String (= Zeichenkette). Für die Weiterarbeit soll das Datenformat in den sogenannten Double (= Kommadezimalzahl) umge-

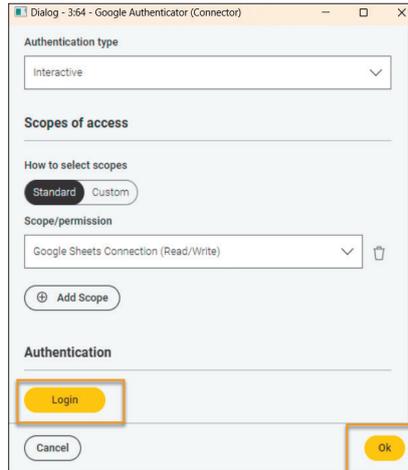


Abb. 3: Log-in zu eigenem Chrome-Browser konfigurieren

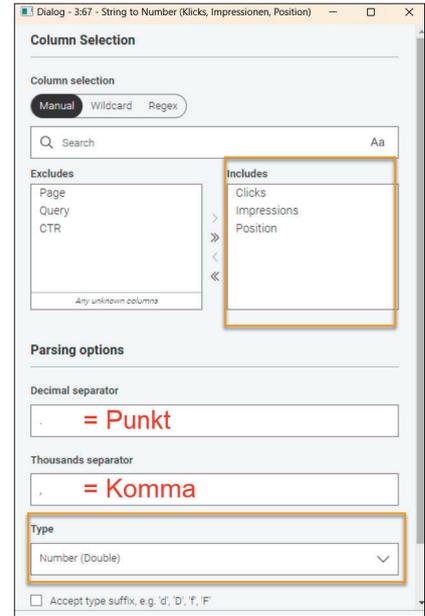


Abb. 4: Umwandlung des Datenformats

**INFO**

Exkurs: Um zu erkennen, in welchem Datenformat Werte in KNIME dargestellt sind, gibt es eine einfache Ansicht: Ist ein Knoten ausgeführt → Rechtsklick auf den Knoten → Auswahl „Transformed input“ (kleines Lupensymbol). Vor allen Spaltennamen wird in der Vorschau ein Buchstabe angezeigt, der für das entsprechende Datenformat steht. Zum Beispiel: S = String, I = Integer, D = Double. In Abbildung 5 ist zu sehen, wie sich die Symbole durch die Umwandlung verändern. Damit kann auch immer überprüft werden, ob alle Werte wie gewünscht geändert wurden (Abbildung 5).

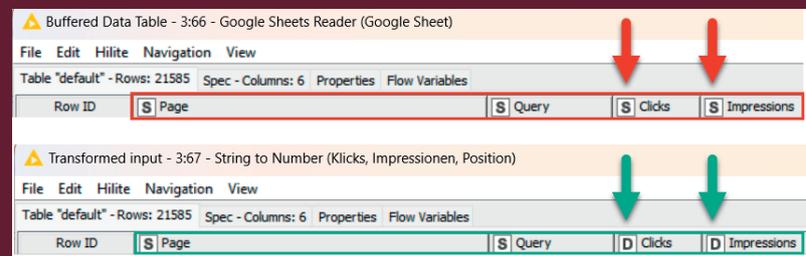


Abb. 5: Ergebnis der Umwandlung des Datenformats

wandelt werden. Ansonsten ist in nachfolgenden Schritten keine Berechnung der Werte möglich. Zur Umwandlung wird der Knoten STRING-TO-NUMBER verwendet. Alle Werte, die umgewandelt werden sollen, werden in „Includes“ hinzugefügt. Zusätzlich werden die Separatoren (= Trennzeichen) der Werte definiert (Abbildung 4). Hinweis: Google Sheets verwendet die englischen Trennzeichen, das heißt, der Punkt ist das Dezimaltrennzeichen (im Deutschen das Komma) und das Komma ist das Tausendertrennzeichen (im Deutschen der Punkt).

**Schritt vier:** Nun geht es um die eigentliche Aufbereitung. Dabei sollen in diesem Beispiel die fünf meistgeklickten Suchanfragen je URL aus den Daten gefiltert werden. Um diese Datenansicht zu erhalten, ist das sogenannte Looping notwendig. Beim Prozess des Loopings wird ein Vorgang so häufig wiederholt, bis der komplette Datensatz anhand einer bestimmten Vorgabe verarbeitet ist. In diesem Looping-Prozess werden für jede URL des Datensatzes die Top-Suchanfragen aufgelistet. Ist diese Aufgabe abgeschlossen, wird zur nächsten URL gesprungen. Bei zehn URLs läuft das Looping demnach zehn-

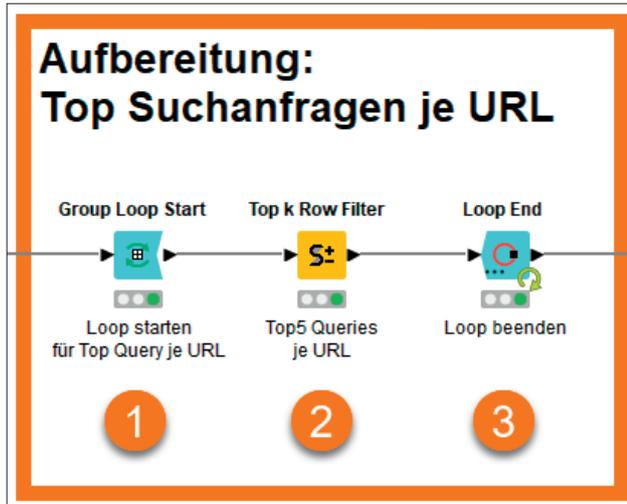


Abb. 6: Aufbereitung der GSC-Daten durch Looping-Prozess

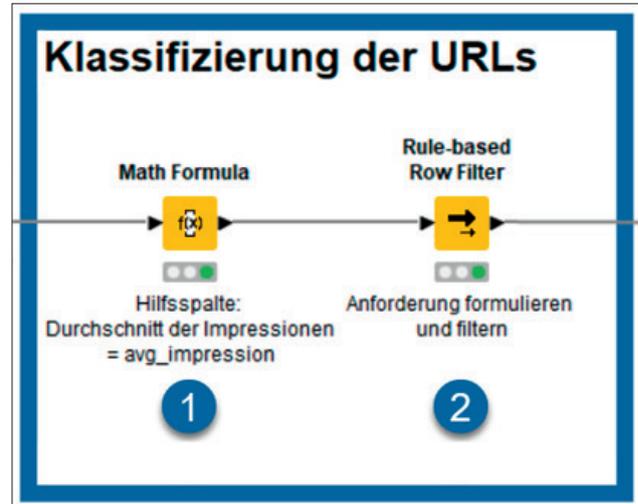


Abb. 7: Klassifikation der Daten in zwei Knoten

## HINWEIS

Es geht hier um die grundlegende Konfiguration der Klassifizierung von URLs. Je nach Anwendungsfall und Anforderung gibt es unendlich viele Möglichkeiten, die URLs zu klassifizieren. Ziel ist es dabei immer, möglichst nur Rankings zu betrachten, die tatsächlich wichtig sind. Ein Ranking mit einer Impression und einem Klick, das zu einer Klickrate von 100 % führt, wäre mit Sicherheit nicht optimierungswürdig, auch wenn das Ranking auf Position vier läge. Der einfachste Weg, um ein solches Rauschen aus den Daten zu filtern, wäre es, alle Rankings mit Impressionen kleiner als zehn aus den Daten zu entfernen. Ein drastischerer Ansatz ist der oben beschriebene: Die Werte der Impressionen müssen mindestens dem Durchschnitt der Impressionen des Datensatzes entsprechen.

mal. Wie in Abbildung 6 gezeigt wird das Looping in KNIME so aufgebaut:

- » Zunächst wird der GROUP-LOOP-START (Abbildung 6, Ziffer 1) benötigt. Dazu wird ausgewählt, für welche Spalte das Looping ausgeführt werden soll. Das ist hier die URL, weshalb in der Konfiguration die Spalte „Page“ ausgewählt wird.
- » Als Nächstes wird die Aufgabe des Loopings definiert. Das ist hier das Ranken der fünf meistgeklickten Suchanfragen je URL. Dafür gibt es den TOP-K-ROW-FILTER (Abbil-

dung 6, Ziffer 2). In der Konfiguration wird die Anzahl der Zeilen auf „5“ gesetzt und die Art der Sortierung auf „Sort by Clicks“ und „Descending“ (= absteigend) eingestellt.

- » Der dritte Knoten beendet das Looping und rahmt die Anforderung an den Loop ein. Im Knoten LOOP-END kann die Default-Konfiguration beibehalten werden. Mit Rechtsklick → „Execute“ wird der Loop ausgeführt.

Hierbei wird nun sichtbar, dass die drei Knoten immer wieder ausgeführt werden, bis alle einzigartigen URLs durchlaufen sind. Bei größeren Datensätzen kann das ein wenig mehr Zeit in Anspruch nehmen.

## Fokus auf das Wichtige: Klassifikation der Daten in KNIME

Nun geht es um die Klassifizierung der URLs. Ziel ist es dabei, den Datensatz auf die wichtigsten Werte zu minimieren. Denn selten müssen tatsächlich alle URLs eines bestimmten Leistungszeitraums betrachtet werden. Hier sollen folgende Kombinationen aus URL und Suchanfrage gefiltert werden:

- » Die Suchanfrage hat zur zugehörigen URL mehr als den Durchschnitt an Impressionen erzielt und die Suchanfrage eine Position zwischen vier und elf. Dies führt hier zur Definition Low-Hanging-Fruit-Potenzial.

## Schritt fünf

Um die Suchanfragen nach dem Durchschnittswert der Impressionen zu filtern, wird nun eine Berechnung notwendig. Das funktioniert in KNIME mit dem Knoten MATH-FORMULA (Abbildung 7, Ziffer 1). Im Knoten selbst gibt es eine Vielzahl an vorgefertigten Formeln, unter anderem für die Bildung des Durchschnitts. Die korrekte Formel dafür ist:  $COL\_MEAN() = \text{Durchschnitt einer Spalte}$ . In der Konfiguration (Abbildung 8) wird einfach die Formel aus „Functions“ mit Doppelklick ausgewählt. Danach wird in die Klammern der Formel die entsprechende Spalte, die oben links in der Column List gezeigt wird, eingefügt. Hier: Impressions. Mit Auswahl „Append Column“ und einer selbst gewählten Benennung wird mit „OK“ und der Ausführung des Knotens eine neue Spalte an den Datensatz angehängt. Diese neue Spalte heißt hier: „avg\_impression“.

## Schritt sechs

Nun werden die Zeilen des Datensatzes mithilfe des RULE-BASED ROW-FILTERS (Abbildung 7, Ziffer 2) klassifiziert. Eine schöne Einführung mit vielen Beispielen zu den RULE-Knoten ist in Ausgabe #82 von Sarah Zeus zu finden. In der Konfiguration des Knotens (Abbildung 9) wird nun definiert, welche Daten aufgrund der formulierten Regel eingeschlossen oder ausgeschlos-

sen werden sollen. Da in der Expression am Ende der Zeile TRUE steht und die Auswahl auf „Include TRUE matches“ gestellt ist, werden mit Ausführung des Knotens automatisch alle Daten ausgeschlossen, die nicht mit der Regel übereinstimmen.

An dieser Stelle wird noch einmal deutlich, wie individuell die Klassifizierung sein kann. Denn: Befinden sich im minimierten Datensatz kaum oder keine Daten, kann die Konfiguration des Knotens angepasst werden. Sollen im RULE-BASED ROW-FILTER weitere Anforderungen gemacht werden, können diese ganz einfach in weiteren Zeilen ergänzt werden. Entscheidend ist in diesem Knoten immer, dass für jede Regel am Ende der Zeile definiert wird, ob sie bei Übereinstimmung die Daten einschließen (= TRUE) oder ausschließen (= FALSE) soll.

**Schritt sieben**

Wie in Schritt zwei werden als Knoten der GOOGLE-ATHENTICATOR und der GOOGLE-SHEETS-CONNECTOR benötigt und wie zuvor konfiguriert. Um die Daten nun tatsächlich zu exportieren, wird der Knoten GOOGLE-SHEETS-WRITER gebraucht. Hinweis: Wie in Abbildung 10 zu sehen, wird im WRITER-Knoten an den oberen Port der ATHENTICATOR und der CONNECTOR angebunden. An den unteren Port des WRITER-Knotens wird der vorhergehende Workflow angebunden.

In der Konfiguration des WRITER-Knotens muss ein Name für das neu zu erstellende Google Sheet angegeben werden, ansonsten kann die Konfiguration mit den Default-Einstellungen übernommen werden. Durch die Ausführung des Knotens wird aus den Daten ein neues Google Sheet erstellt und der Workflow ist abgeschlossen.

**Fazit**

Der vorgestellte Workflow zeigt, dass KNIME nicht nur zur Auswertung

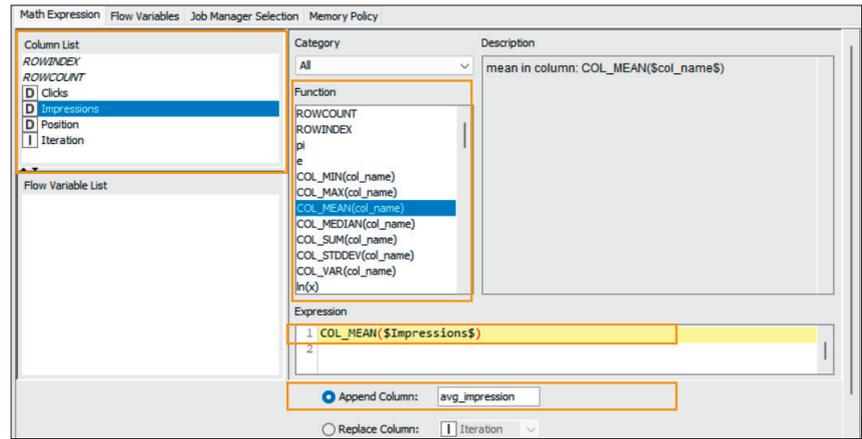


Abb. 8: Berechnung des Durchschnitts der Impressionen des Datensatzes

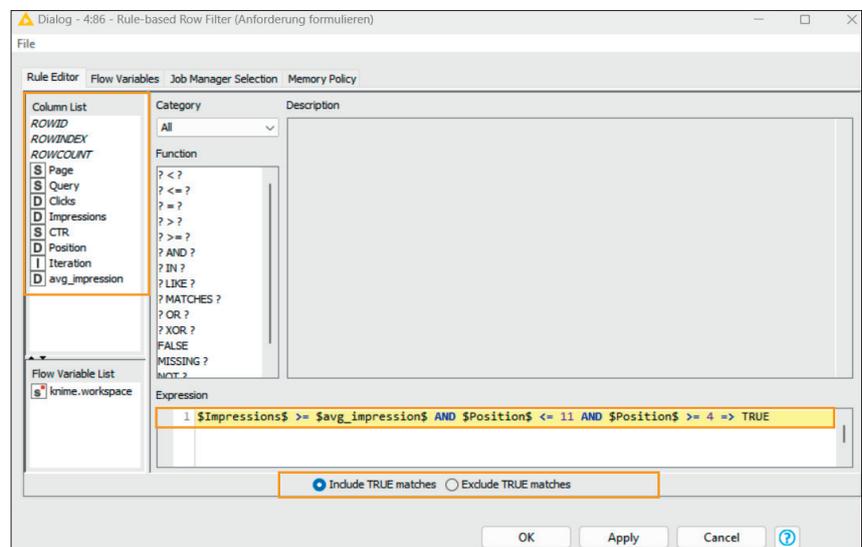


Abb. 9: Definition der Anforderung zur Klassifizierung der Daten

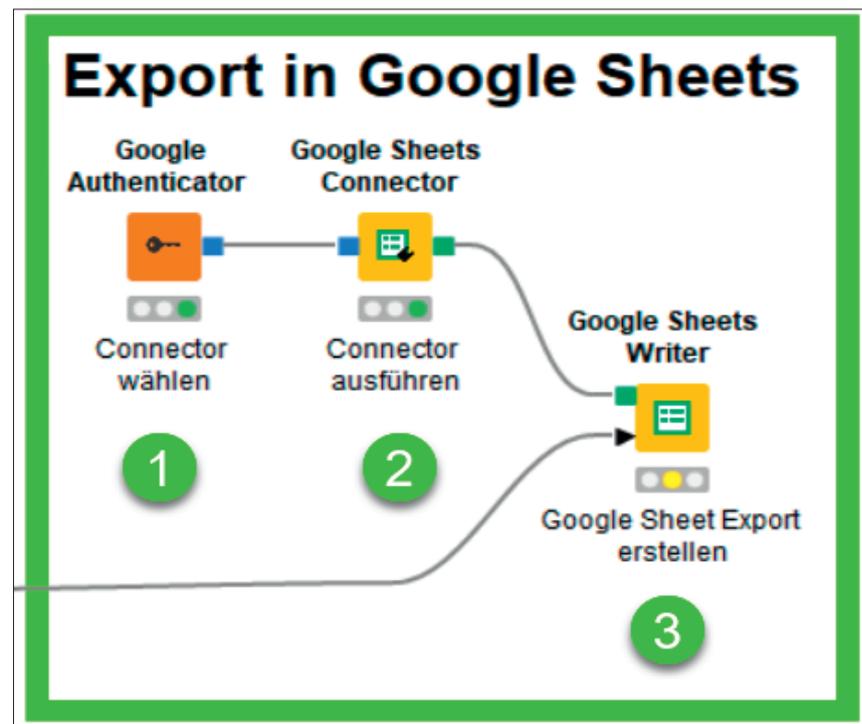


Abb. 10: Export der Daten als Google Sheet

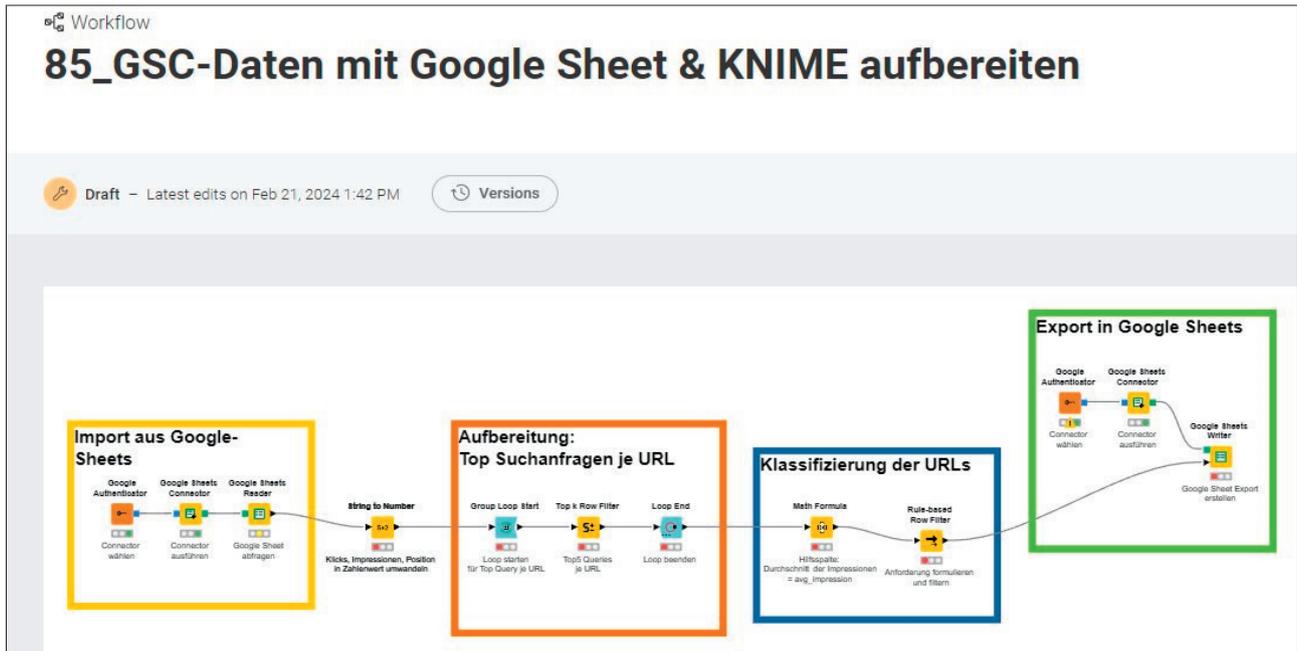


Abb. 11: Der fertige Workflow ist unter [einfach.st/knime932](https://einfach.st/knime932) abrufbar.

von Daten geeignet ist, sondern es auch ermöglicht, Daten für weitere Auswertungen vorzubereiten und in die gewünschte Form zu bringen. In den einzelnen Konfigurationen wird schnell deutlich, wie vielfältig das Tool genutzt werden kann und wie leicht die Anpassung des Workflows zugunsten der

eigenen Daten möglich ist. Das Prinzip lautet hier „Trail and Error“. Denn Daten der gleichen Struktur können zwar immer wieder durch diesen Workflow fließen, einzelne Parameter müssen jedoch je nach Betrachtung immer wieder neu definiert werden, um zum gewünschten Ziel zu kommen.

Der Workflow mit allen hier vorgestellten Konfigurationen ist im KNIME-Hub unter [einfach.st/knime932](https://einfach.st/knime932) abrufbar und kann direkt in die eigene KNIME-Umgebung eingelesen werden. Viel Spaß beim Ausprobieren!