

Michael Göpfert

# DO IT:

## Eine einfache Analyse von Search-Console-Daten mit KNIME

Verschiedene Datensätze aus drei Quellen sollen zusammen geführt werden. Leider ist das meist nicht so einfach, da z. B im Datensatz A das Datum im Format „2021.01.04“ hinterlegt ist, in den beiden anderen als „4. Jan. 2021“. Die URL im Datensatz B ist im Muster „www.meinshop.de/index.html“ hinterlegt, in Datensatz C als „https://www.meinshop.de/index.html“ und im Datensatz A gar nur mit „/index.html“. Datensatz B enthält statt deutschem Umlauten grumpfelige Sonderzeichen. Kennen Sie das? Meist bleibt einem nichts anderes übrig, als alle drei Datensätze vor der Zusammenführung manuell via Excel entsprechend mit Formeln herzurichten. Eine mühselige Arbeit und eigentlich unnötig – wenn man ganz einfach KNIME einsetzen würde. Denn alles was Sie dort einmal als Regelwerk hinterlegt haben, brauchen Sie später nie wieder tun. Nächsten Monat jagen Sie dann einfach die drei Datensätze durch KNIME und alle Umformungen plus die Zusammenführung geht ab jetzt vollautomatisch und in Echtzeit.

Bereits in Ausgabe 53 wurde das Thema KNIME aufgegriffen und ein Workflow zur Analyse eines Screaming-Frog-Crawls gezeigt. Nach über zwei Jahren ist es an der Zeit, das Thema Einsteigern und Fortgeschrittenen noch einmal näher zu bringen.

Dieses Mal soll es allerdings darum gehen, wie man mit wenigen Handgriffen eine Analyse der Search-Console-Daten zu ermöglichen. Experte Michael Göpfert zeigt, wie das funktioniert. Da KNIME eine freie Software ist, können Sie diese kostenlos downloaden und am besten gleich step-by-step direkt alles am Rechner nachvollziehen. Und auch wenn Sie die Daten von Google für Ihre Domain aus der Search Console nicht interessieren sollten, können Sie durch Übertragung auf ein anderes Datenproblem daraus Wissen ziehen. Wichtig ist, dass Sie sich einfach mal hinsetzen und ES MACHEN...

### DER AUTOR



**Michael Göpfert** ist Senior SEO Consultant bei der diva-e Advertising GmbH. KNIME ist für ihn ein unverzichtbares Werkzeug für Analyse von Keyword-Daten und Crawls.

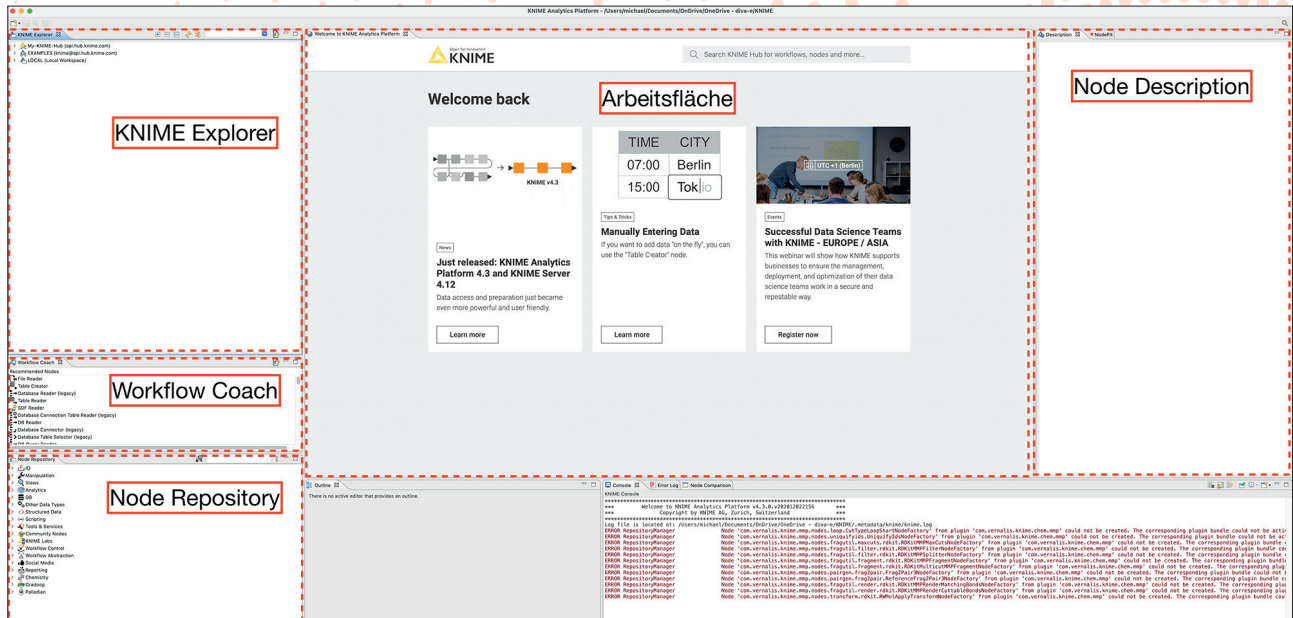


Abb. 1: Die Benutzeroberfläche von KNIME

## Was ist KNIME?

KNIME steht für „Konstanz Information Miner“ und ist eine freie Software für Windows, Mac und Linux, die an der Universität Konstanz entwickelt wurde, um große Datenmengen zu analysieren und zu verarbeiten. Beispielsweise:

- » Crawl-Analyse: Repetitive Aufgaben wie das Analysieren von Crawls lassen sich automatisieren, indem ein Workflow einmalig aufgesetzt wird und beliebig oft mit neuen Crawl-Daten ausgeführt werden kann.
- » Data Mining und Aufdecken von Mustern und Zusammenhängen in großen Datenmengen. Selbst sehr große Datenmengen können performant verarbeitet werden.
- » Auswertung von Performance-Daten: Segmentierung, Analyse und Gruppierung von Keywords und deren Performance.

Dabei setzt KNIME auf eine intuitive grafische Benutzeroberfläche und kann ohne Programmierkenntnisse genutzt werden. Stattdessen werden bei KNIME sogenannte Workflows erstellt, die aus einzelnen Modulen (sog. „Nodes“) bestehen, von denen jedes eine eigene Aufgabe erfüllt. Die Nodes werden miteinander verbunden und durch Ausführen des Workflows nacheinander abgearbeitet.

Ein einfacher Workflow könnte beispielsweise wie folgt aufgebaut sein:

1. Node: Importieren der Daten
2. Node: Filtern der Daten
3. Node: Visualisieren der Daten

Die Nodes werden von KNIME zur Verfügung gestellt oder können kostenfrei über die Website [nodepit.com](http://nodepit.com) heruntergeladen werden.

## Die Benutzeroberfläche

Nach der Installation fragt KNIME ([www.knime.com](http://www.knime.com)) Sie nach einem Verzeichnis, in dem Ihre Workflows abgespeichert werden sollen („Workspace“). Anschließend werden die Benutzeroberfläche und eine Willkommenseite angezeigt.

Vor der Erstellung Ihrer ersten Workflows gilt es, sich mit der Benutzeroberfläche vertraut zu machen: Auf der linken Seite befinden sich in der Standardansicht drei Fenster:

- » KNIME Explorer: Hier finden Sie alle Workflows, die im Workspace abgespeichert wurden.
- » Workflow Coach: Der Workflow Coach macht Vorschläge, welche Node als Nächstes verwendet werden könnte. Dies funktioniert allerdings nur, wenn Sie bei der Installation von KNIME zugestimmt haben, Nutzerstatistiken an KNIME zu übermitteln.

## WAS MACHT KNIME EIGENTLICH?

KNIME ändert i. W. Datensätze nach vorgebbaren Regeln. „Lösche alle Zeilen, bei denen in Spalte 5 kein Wert steht“. Oder auch „Füge in alle leeren Zellen in Spalte 11 den Wert 0 ein“. „Füge bei allen Inhalten in Spalte 2 vorne ein >https://< ein“. „Prüfe, ob in Spalte 2 tatsächlich nur numerische Daten eingetragen sind“. „Ändere den Zeichensatz der Daten von ISO-8859-1 auf UTF-8“. All das sind typische Aufgaben, die man mit KNIME in einem Rutsch erledigen kann. Aber auch Filter („Füge nur die Daten aus Datensatz A, bei denen in Spalte 8 ein Wert größer 100 € steht, dem Datensatz B hinzu“) und weitere Datenmodifikationen lassen sich relativ einfach hinterlegen. Der Vorteil liegt auf der Hand. Man definiert nur einmal alles, was geändert werden muss, wie die Daten am Ende zusammengebaut werden und kann das fortan immer und immer wieder nutzen. Wer häufiger mit Daten aus verschiedenen Quellen z. B. für Reportings umgehen muss, für den ist KNIME eine wahre Zeitspar-Goldgrube! Und natürlich kann KNIME noch sehr viel mehr. Aber für den Einstieg soll diese Erklärung erst einmal reichen.

» Node Repository: Hier befinden sich alle installierten Nodes. Durch die Suchfunktion lässt sich die große Menge an Nodes leicht beherrschen. Da die Nodes nicht immer logisch kategorisiert sind, ist es sinnvoll, anhand der Suchfunktion nach der benötigten Funktion zu suchen (z. B. „Filter“).

In der Mitte befindet sich noch die Willkommenseite von KNIME. Sobald ein neuer Workflow erstellt wird (Menü » New » New Workflow), erscheint an dieser Stelle die Arbeitsfläche („Workflow Editor“). Rechts davon befindet sich noch der Tab „Description“. Dieser zeigt Ihnen zu jeder ausgewählten Node eine Beschreibung an. Wird beispielsweise im Node Repository nach einer Funktion gesucht und eine potenziell infrage kommende Node ausgewählt, können Sie anhand der Description beurteilen, ob diese Node die Anforderungen erfüllt.

Alle weiteren Fenster wie Outline oder Console sind für dieses Beispiel zunächst nicht relevant und werden deshalb nicht weiter behandelt.

### Der erste Workflow

Wird wie oben beschrieben über Menü » New » New Workflow ein neuer Workflow erstellt und dieser im Workspace gespeichert, zeigt sich eine leere karierte Fläche, die es nun mit Leben zu füllen gilt.

### Vorbereitung: Daten aus der Google Search Console via Google Sheets exportieren

Als Beispiel-Workflow soll eine Auswertung der Keywords aus der Google Search Console (GSC) dienen. Um die Search-Console-Daten in KNIME importieren zu können, müssen Sie zunächst eine Tabelle mit den GSC-Daten erstellen.

Um die Begrenzung des Exports von GSC-Daten auf 1.000 Zeilen zu umgehen, können Sie die Keyword-Daten mit

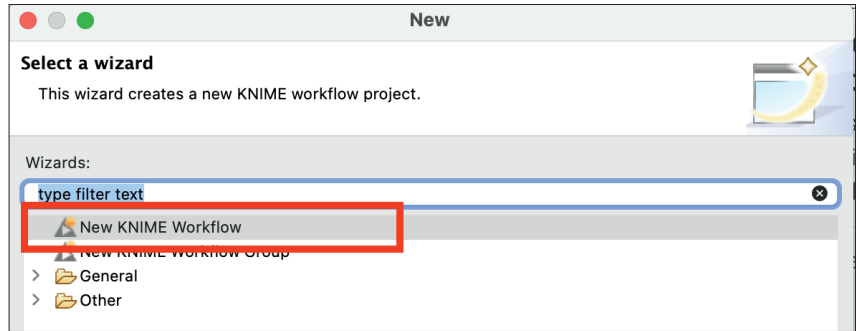


Abb. 2: Der Workflow Wizard ermöglicht das Erstellen eines neuen Workflows

dem kostenlosen Google-Sheets-Add-on Search Analytics for Sheets direkt aus der Search Console in Google Sheets importieren. Dazu rufen Sie in Google Sheets den Menüpunkt Add-ons auf, im Untermenüpunkt Add-ons aufrufen führt eine Suche nach Search Analytics for Sheets direkt zum Add-on und dessen Installation. Bei der Installation müssen Sie dem Add-on noch Zugriff auf die Search Console gewähren. Anschließend kann das Add-on im Menü unter Add-ons » Search Analytics for Sheets » open Sidebar geöffnet werden.

Nach Aufrufen des Add-ons wählen Sie eine Website aus der Search Console und einen Zeitraum aus. Die Daten werden im Feld Group By nach Query gruppiert, zusätzlich empfiehlt es sich, ein paar Filter (Country, Device und ein bestimmtes Verzeichnis) zu setzen, um möglichst aussagekräftige Daten zu erhalten. Anschließend lädt das Add-on die Daten in Google Sheets.

Nachdem die Daten in Google Sheets geladen wurden, laden Sie die Datei als Excel-Datei auf die lokale Festplatte herunter.

Übrigens: Andere Dateiformate wie CSV sind für KNIME zwar auch kein Problem, bei den GSC-Daten im Excel-Format müssen jedoch die wenigsten Anpassungen vorgenommen werden.

### GSC-Daten importieren

Liegen die Daten auf Ihrem Rechner, beginnt der Import in KNIME. Dazu suchen Sie im Node Repository nach „Excel“ und ziehen die Node Excel Reader per Drag & Drop (alternativ: Doppelklick) in die karierte Arbeitsfläche.

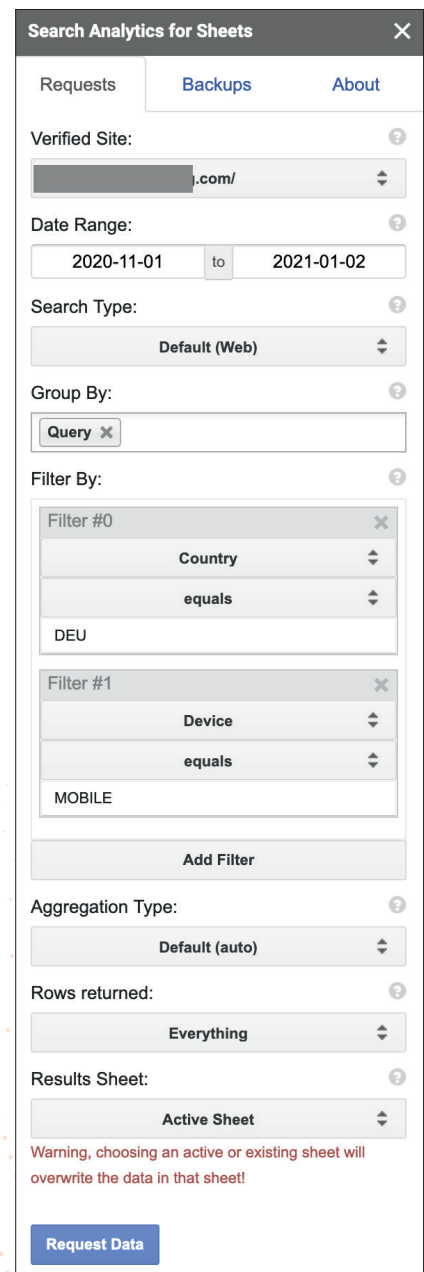


Abb. 3: Datenimport in Google Sheets via Search Analytics for Sheets

Die Node zeigt nun einen roten Punkt und ein warnendes Dreieck an. Das signalisiert, dass sie in ihrem jet-

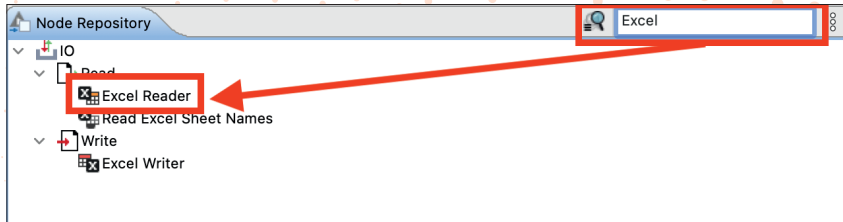


Abb. 4: Eine Suche nach „Excel“ zeigt die Nodes für Excel-Dateien im Node Repository an; mit der Node „Excel Reader“ lassen sich Excel-Dateien in KNIME importieren

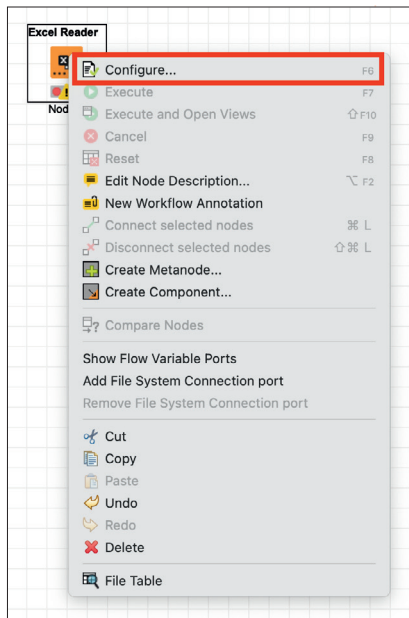


Abb. 5: Kontextmenü der Excel Reader Node

Row ID	S Query	I Clicks	I Impre...	D CTR	D Position
Row0	helmlampe mtb	169	1869	0.09	3.499
Row1	five ten winter...	101	466	0.217	1.004
Row2	five ten freeri...	51	556	0.092	2.755
Row3	endura singlet...	49	427	0.115	3.087
Row4	endura singlet...	49	485	0.101	3.06
Row5	mtb helmlampe	43	475	0.091	4.678
Row6	garmin fenix ...	39	339	0.115	2.72
Row7	evoc fr trail	37	1294	0.029	4.094
Row8	mtb helmlam...	30	172	0.174	3.634
Row9	five ten freeri...	29	271	0.107	4.092
Row10	mtb trails bay...	29	160	0.181	1.944
Row11	helmlampe mt...	26	90	0.289	1.178
Row12	five ten eps	25	214	0.117	2.505
Row13	garmin fenix ...	25	80	0.312	1
Row14	fenix 6 naviga...	23	89	0.258	1.685
Row15	fenix 6s pro test	23	137	0.168	2.635
Row16	five ten winter...	23	141	0.163	2.348
Row17	garmin fenix ...	21	66	0.318	1.182
Row18	mtb schienbei...	21	76	0.276	1.461
Row19	federweg	19	253	0.075	1.451
Row20	five ten wasse...	19	128	0.148	2.031
Row21	garmin fenix ...	19	49	0.388	3.449
Row22	karwendelrun...	19	136	0.14	4.184
Row23	mtb helmlam...	19	146	0.13	3.534
Row24	mtb im winter	19	326	0.058	6.911
Row25	mtb hose lang...	18	166	0.108	6.886
Row26	five ten winter	17	102	0.167	1.147
Row27	garmin fenix ...	17	48	0.354	1
Row28	garmin fenix ...	16	558	0.029	8.996
Row29	endura singlet...	15	101	0.149	5.149
Row30	evoc hip pack	15	1163	0.013	5.994
Row31	fenix 6 pro test	15	217	0.069	6.756
Row32	garmin fenix ...	15	81	0.185	3.432
Row33	garmin fenix ...	15	498	0.03	9.789
Row34	garmin fenix ...	15	272	0.055	2.849

Abb. 6: Die aus der Excel-Datei importierten Daten

zigen Zustand noch nicht ausgeführt werden kann und zunächst von Ihnen konfiguriert werden muss.

Mit einem Rechtsklick auf die Node öffnet sich ein Kontextmenü, dessen oberster Punkt Configure... lautet. Durch Anklicken öffnet sich ein Fenster, in dem Sie die gewünschten Einstellungen vornehmen können.

In diesem Fall beschränkt sich das Konfigurieren auf das Auswählen der Excel-Datei, die aus Google Sheets heruntergeladen wurde. Dazu wird über den Button Browse die zu importierende Excel-Datei ausgewählt. Alle anderen Einstellungen erkennt KNIME automatisch.

Im unteren Teil des Konfigurationsfensters sehen Sie eine Vorschau der zu importierenden Daten. Wenn diese Vorschau Ihren Erwartungen entspricht, können Sie durch Klicken der OK-Schaltfläche die Einstellung übernehmen und

die Node ändert ihren Zustand in „ausführbar“. Das rote Dreieck ist nun verschwunden und die „Ampel“ ist nicht mehr rot, sondern gelb.

Die Node kann nun über das Kontextmenü (Execute) oder die Menüleiste (grüner Play-Button) ausgeführt werden. Durch einen Rechtsklick auf die Node und die Auswahl File Table werden die importierten Daten angezeigt.

Diese Art zu arbeiten ist etwas gewöhnungsbedürftig, weil Sie die Daten, mit denen Sie arbeiten, nicht sofort sehen. Die Möglichkeit, die Daten über den untersten Punkt des Kontextmenüs aufzurufen, gibt es aber bei fast allen Nodes, sodass Sie stets überprüfen können, ob die getroffenen Einstellungen die richtigen waren. Ist dies nicht der Fall, können die Einstellungen jederzeit geändert werden. Für das Wirksamwerden muss die Node erneut ausgeführt werden.

## Datentypen

Bevor mit den importierten Daten gearbeitet wird, sollten Sie überprüfen, ob diese im richtigen Datentyp importiert wurden. Wie in Excel können beispielsweise Berechnungen nur dann durchgeführt werden, wenn die Spalte als „Zahl“ formatiert ist. In KNIME gibt es mehrere Datentypen für Zahlen, beispielsweise Double (für Dezimalzahlen) und Integer (für Ganzzahlen). In diesem Fall steht hier im Spaltentitel „Query“ ein S (für String), bei den Spalten „Clicks“ und „Impressions“ das I für die Ganzzahl und bei der „CTR“ und der „Position“ das D für die Dezimalzahl. Es besteht also kein Handlungsbedarf.

## Troubleshooting Datentyp

Importieren Sie Daten aus einer anderen Quelle, kann es sein, dass Sie die Datentypen anpassen müssen. Dazu gibt es die Node String to Number, falls Zahlen fälschlicherweise als String importiert wurden. Auch wichtig zu wissen ist, dass KNIME ausschließlich das amerikanische Datenformat beherrscht, d. h., das Tausendertrennzeichen ist kein Punkt, sondern ein Komma, umgekehrt werden Dezimalzahlen mit einem Punkt getrennt.

## Aufbereiten der GSC-Daten

Was bei den importierten Daten auffällt: Neben der möglichen Anpassung des Datentyps kann es auch passieren, dass die Daten selbst angepasst werden müssen. Im aktuellen Beispiel fällt beim Vergleich der Originaldaten mit den importierten Daten auf: Die Click-Through-Rate-Werte (CTR) sind um den Faktor 100 zu gering und müssen erst noch angepasst werden. Das ermöglicht die Node Math Formula. Nachdem Sie diese per Drag & Drop auf die Arbeitsfläche gezogen haben, verbinden Sie das schwarze Dreieck (Output Port) der Excel Reader Node mit dem linken Dreieck (Input Port) der Math Formula Node per Drag & Drop.

Durch das Verbinden der Nodes kann die Math Formula nun auf die vom Excel Reader importierten Daten zugreifen.

Auch die Math Formula Node muss zunächst konfiguriert werden. Beim Öffnen des Konfigurationsfensters der Node zeigt sich, dass links oben die aus Excel importierten numerischen Spalten aufgeführt sind. Die Spalte mit den Suchanfragen taucht hier nicht auf, weil es sich dabei nicht um einen numerischen Datentyp handelt. Nun haben Sie zwei Möglichkeiten: Es kann im Bereich Function eine Funktion ausgewählt oder im Feld Expression eine eigene Funktion eingegeben werden. Da lediglich die Spalte CTR mit 100 multipliziert werden soll, reicht es, wenn Sie die Spalte CTR per Doppelklick in das Expression-Feld übertragen und dahinter \*100 schreiben:  $\$CTR\$*100$ .

Zusätzlich besteht die Möglichkeit, dass Sie das Ergebnis der Berechnung in eine neue Spalte schreiben (Append Column) oder die bestehende Spalte überschreiben (Replace Column).

Manipulierte Daten in eine neue Spalte zu schreiben, ist oft sehr nützlich, wie Sie in einem späteren Beispiel sehen werden. Für diesen Vorgang reicht es jedoch, die Spalte CTR mit den neu berechneten Daten zu überschreiben.

Nachdem Sie die Node ausgeführt haben, steht in der Spalte CTR der neue Wert.

Damit ist Ihre erste kleine Automatisierung abgeschlossen, denn auch wenn Sie ein neues File mit GSC-Daten importieren, kann der aufgebaute Workflow erneut genutzt werden. Wichtig ist dabei nur, dass die Namen und Datentypen der Spalten identisch sind, da die Nodes auf diese zugreifen.

### Statistische Auswertungen der Daten

Nachdem Sie alle Daten in das richtige Format gebracht haben, ist es an der Zeit, erste Erkenntnisse zu gewinnen. Eine einfache und zugleich

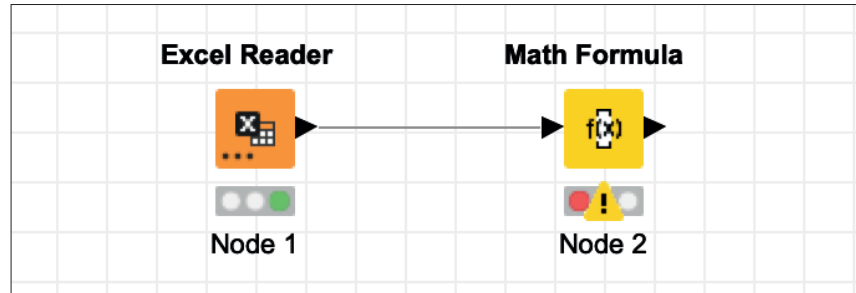


Abb. 7: Durch Verbinden der Nodes kann die Node Math Formula auf die via Excel Reader importierten Daten zugreifen

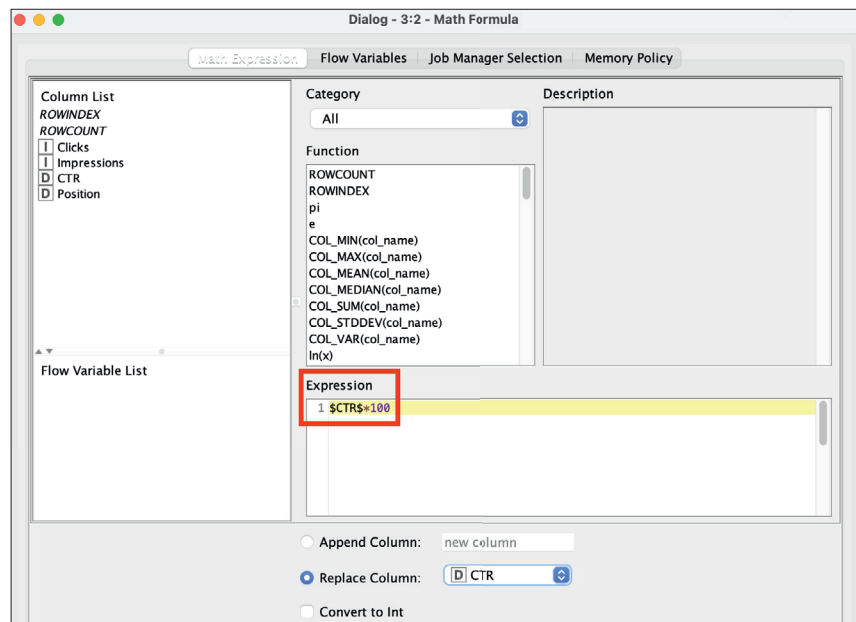


Abb. 8: Multiplikation in der Node „Math Formula“

sehr eindrucksvolle Node ist die Node Statistics. Nachdem diese auf die Arbeitsfläche gezogen und mit der Node Math Formula verbunden wurde, kann sie über das Kontextmenü in den Konfigurationseinstellungen angepasst werden. Zunächst setzen Sie den Haken bei Calculate median values.

Darunter befinden sich zwei Felder, in denen die Spalten aufgeführt sind, die beim Ausführen der Node berücksichtigt (Include) oder ignoriert (Exclude) werden. Um alle numerischen Spalten auszuwerten, müssen Sie die Spalten CTR und Position auswählen und durch Klicken des nach rechts zeigenden Pfeils in das Include-Feld übertragen. Umgekehrt wird die Spalte Query durch Auswahl und Drücken des nach links zeigenden Pfeils exkludiert.

Nach Klicken des OK-Buttons ist die Node bereit, ausgeführt zu werden.

Um sich eine Visualisierung anzeigen zu lassen, wählen Sie im Kontextmenü Execute and open views. Daraufhin öffnet sich ein Fenster mit tabellarischen und grafischen Auswertungen. Für jede Spalte werden unter anderem der höchste, der niedrigste, sowie der Median und der Durchschnittswert angezeigt.

Anhand der Verteilung der Daten im Histogramm können Sie erkennen, wo sich der Löwenanteil der Keywords in Bezug auf Impressions, Klicks, Rankings und CTR befindet. In diesem Beispiel wird sichtbar, dass nur eine kleine Anzahl an Keywords auch Klicks bringt, während der Großteil der Suchbegriffe nicht zum Traffic der Website beiträgt.

Die erste Analyse in KNIME ist damit abgeschlossen. Um noch ein Stück weiterzugehen, sollen die importierten Keywords nach Brand und

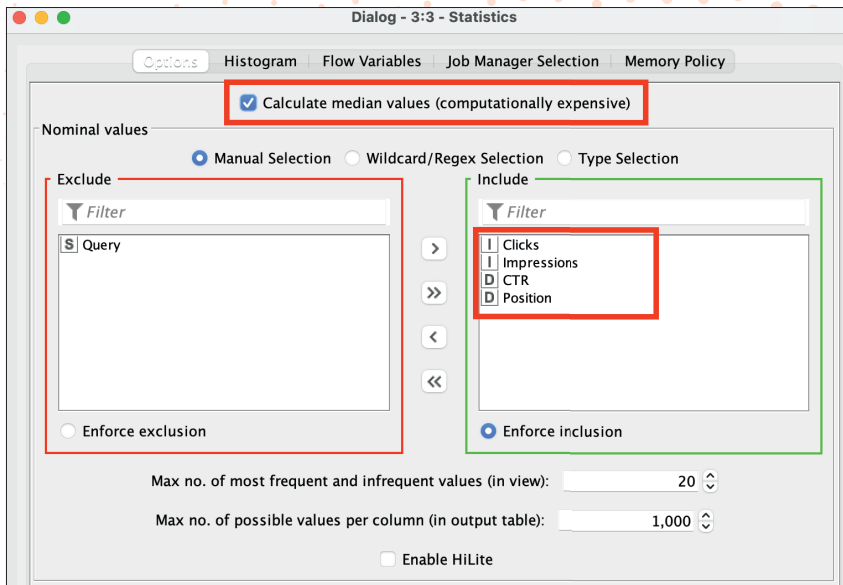


Abb. 9: Fertig konfigurierte Statistics Node

Non-Brand unterschieden werden und die statistische Auswertung für beide Cluster durchgeführt werden.

### Segmentierung nach Brand- und Non-Brand-Keywords

Für das Segmentieren von Daten bietet sich die Node Rule Engine an. Die Rule Engine ist eine sehr mächtige Node, weil sie es Ihnen ermöglicht, Spalten anhand von Regeln zu überprüfen und je Ergebnis einen Wert in eine neue Spalte zu schreiben (oder eine existierende Spalte zu überschreiben). Genau das richtige Werkzeug also, um die Keywords zu clustern.

Die Rule Engine Node können Sie einfach in den bestehenden Workflow einfügen. Dazu ziehen Sie die Node per Drag & Drop vom Repository auf die Verbindungslinie von Math Formula und Statistics. Sobald sich die Linie rot färbt, kann die Node losgelassen werden und fügt sich zwischen den bestehenden Nodes ein. Hier zeigt sich ein weiterer Vorteil von KNIME: Bestehende Workflows können Sie problemlos modifizieren.

Die Konfiguration erfolgt ebenfalls über das Kontextmenü. Wie schon bei der Math Formula Node stehen Ihnen auch bei der Rule Engine verschiedene Felder zur Konfiguration zur Verfügung. Im Feld links können Sie

auch hier auf die vorhandenen Spalten zugreifen. Durch Doppelklicken auf die Query-Spalte wird im Feld Expressions der Eintrag `$Query$` erzeugt. Im nächsten Schritt übertragen Sie im Feld Function ebenfalls durch einen Doppelklick die Funktion `? Matches ?` in das Function-Feld. Sie ermöglicht das Prüfen der Spalte Query gegen einen regulären Ausdruck. Diesen schreiben Sie anschließend in Anführungszeichen in das Expressions-Feld. In diesem Beispiel sollte dies der Name der Brand sein. Ein einfacher regulärer Ausdruck wäre: `.*Brandname.*`. Damit greift die Regel bei allen Suchbegriffen, die den Namen der Brand beinhalten. Nun tragen Sie nach einem `=>` den Wert ein, der in die neue Spalte geschrieben werden soll: Brand. Die finale Regel lautet also

wie folgt: `$Query$ MATCHES „.*Brandname.*“ => „Brand“`.

Da KNIME die Regeln von oben nach unten abarbeitet, können durch eine neue Zeile mit der Regel `$Query$ MATCHES „.*“ => „Nonbrand“` alle Keywords, die noch nicht von der vorherigen Regel als Brand klassifiziert wurden, als Non-Brand gekennzeichnet werden. Um das Ergebnis anzuzeigen, wird eine neue Spalte mit dem Namen Brand? erzeugt.

Nachdem Sie die Rule Engine ausgeführt haben, befindet sich in der Tabelle eine neue Spalte mit dem Namen Brand?, in der für jedes Keyword aufgeführt ist, ob es ein Brand- oder ein Non-Brand-Keyword ist.

### Daten filtern

Nun können Sie die Zeilen auf Brand und Non-Brand filtern. Dazu müssen Sie die Node Row Filter zwischen der Rule Engine und der Statistics Node platzieren und über das Kontextmenü folgende Konfiguration vornehmen: Column to test: Brand?; Use Pattern Matching: Nonbrand.

Nun können die Filter und Statistics Node erneut ausgeführt werden. Die Visualisierungen der Statistics Node beziehen sich jetzt nur noch auf Non-Brand-Keywords.

Um die statistische Auswertung zusätzlich für Brand-Keywords verfügbar zu machen, können Sie die Nodes Row Filter und Statistics einfach

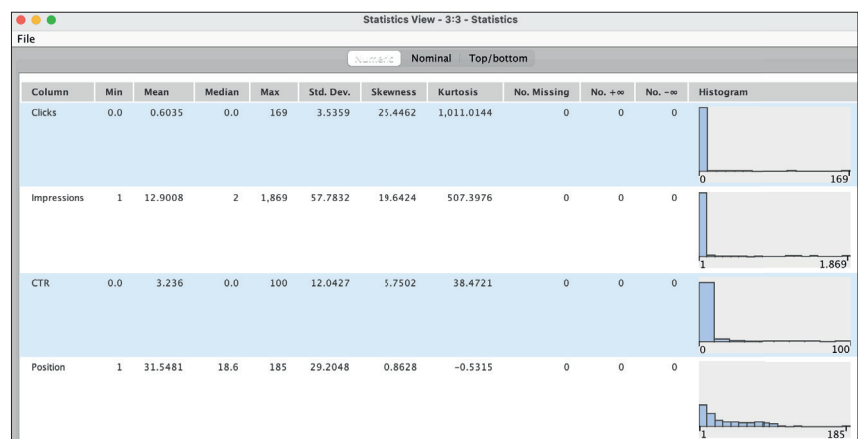


Abb. 10: Auswertung der GSC-Daten mit der Statistics Node

kopieren und einfügen. Anschließend muss nur noch die Rule Engine mit dem eingefügten Row Filter verbunden werden und der Row Filter so konfiguriert werden, dass nur noch Brand-Keywords enthalten sind: Column to test: Brand?; Use Pattern Matching: Brand.

### Bonus: Brand vs. Non-Brand Chart

Nachdem Sie die Daten schon so weit aufbereitet haben, ist die Auswertung der Klick-Anteile nach Brand und Non-Brand nur noch eine Node entfernt: Mit der Node Pie/Donut Chart können Sie die verarbeiteten Daten schnell und einfach in einem Tortendiagramm visualisieren. Sie muss lediglich mit der Node Rule Engine verbunden werden. Anschließend sind in der Konfiguration der Pie Chart Node folgende Einstellungen nötig: Category Column: Brand?; Aggregation Method: Sum, Frequency Column: Clicks. Führen Sie die Node nun über das Kontextmenü und Execute and open Views aus, zeigt sich die Verteilung der Klicks auf Brand und Non-Brand in einem Tortendiagramm.

### Fazit

Sind die Daten erst einmal in KNIME importiert und die ersten Berührungspunkte abgebaut, sind die Möglichkeiten zur Datenanalyse schier grenzenlos. Durch die große Menge an Nodes lässt sich für fast jedes Problem eine Lösung finden. Eine Segmentierung von Keywords in Brand und Non-Brand kratzt da nur an der Oberfläche. Dank des modularen Aufbaus der Workflows kann KNIME auch ganz leicht per „Try and Error“ erlernt werden. ¶

Den fertigen Workflow können Sie hier direkt herunterladen: <https://knime/w/4PQyxSpgME3sJ0oK>

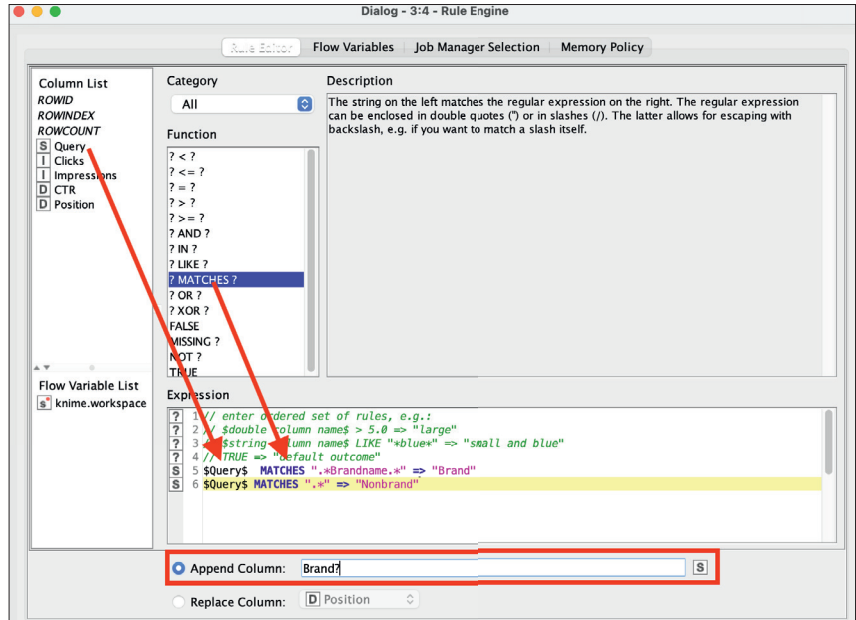


Abb. 11: Konfiguration der Rule Engine, mit der Keywords in Brand und Non-Brand klassifiziert werden

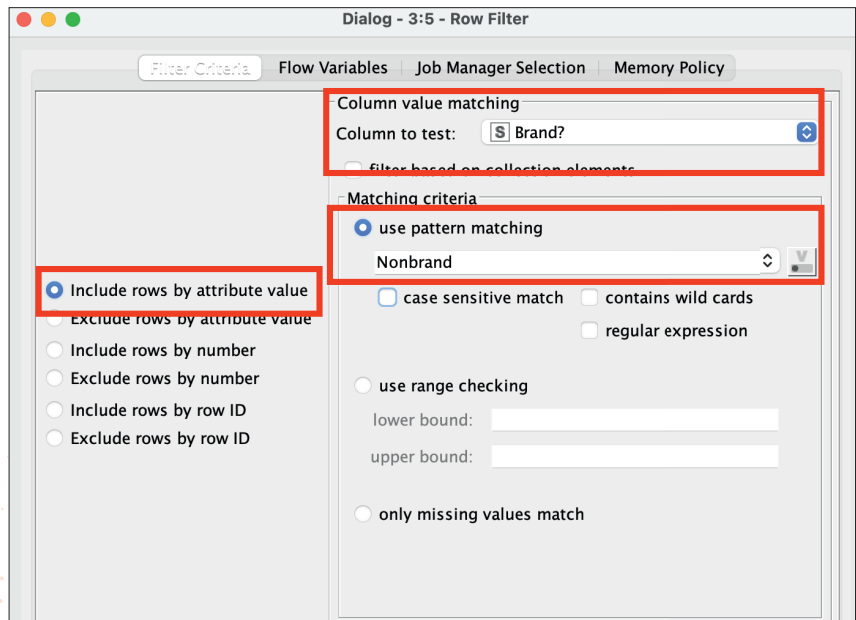


Abb. 12: Konfiguration des Row Filters, sodass nur noch Non-Brand-Keywords in der Tabelle enthalten sind

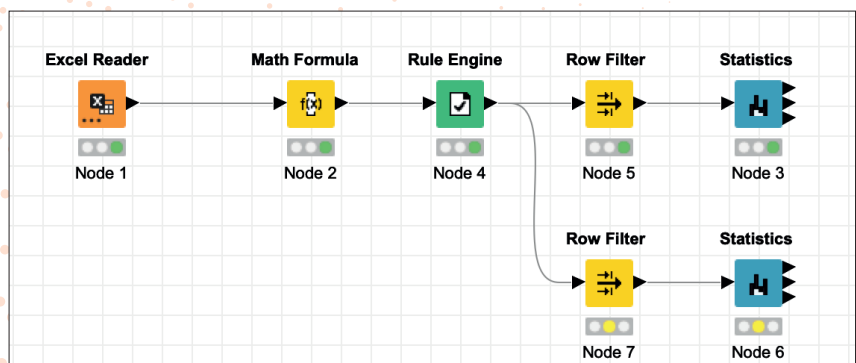


Abb. 13: Erweiterung des Workflows um eine zusätzliche Auswertung für Brand- und Non-Brand Keywords