

Michael Hohenleitner

Search Console Audit mit KNIME

Die Search Console liefert SEOs wichtige Informationen über Traffic und die Keywords, für die eine Website gefunden wird. Doch das Webinterface bietet nur einen unvollständigen Einblick in die Daten. In diesem Artikel lernen Sie, mithilfe von KNIME mehr aus den Search-Console-Daten herauszuholen und wie Sie sich einen schnellen Überblick über die Google-Performance einer Website verschaffen. So bekommen Sie wichtige Daten für die Kontrolle und die Steuerung Ihres SEO-Erfolgs!

Es ist nicht leicht, in den Unmengen an Daten, die selbst bei kleineren Websites anfallen, den Überblick zu behalten. Wenn Sie regelmäßig vor der Herausforderung stehen, Traffic-Daten aus der Google Search Console auszuwerten, sollten Sie sich nicht damit zufriedengeben, durch das Interface zu klicken und zu hoffen, dass Ihnen etwas auffällt, das von Nutzen sein könnte.

Vielmehr sollten Sie auf einen standardisierten Workflow setzen, der Ihnen mit minimalem Aufwand genug Informationen liefert, um ein Bild vom Leistungsvermögen der Website zu haben, Potenziale zu erkennen und Auswertungen zu bekommen, die Sie an Kunden oder Vorgesetzte weitergeben können.

Das Problem mit dem Search Console Interface

Das Webinterface der Search Console lässt sich zwar leicht bedienen, macht aber einige Daten schwer zugänglich oder vorenthält sie komplett. Aufgrund der Limitierung der Daten auf 1.000 Zeilen ist es beispielsweise bei Websites mit viel Traffic oft nicht möglich, Keywords zu finden, für die es zwar Impressionen, aber keine Klicks gibt.

Wünschenswert wäre es also, einen möglichst unverfälschten Rohdatenexport zu erhalten und diesen als Basis für eine Analyse zu verwenden. Um an möglichst vollständige GSC-Daten zu kommen, gibt es verschiedene Wege:

- » Nutzung der API (Programmierschnittstelle): Dazu ist allerdings die Entwicklung einer

DER AUTOR



Michael Hohenleitner arbeitet gerne mit Rohdaten, um diese in maßgeschneiderten Analysen für seine Kunden aufzubereiten.

Foto: zhev / gettyimages.de

country	device	page	query	clicks	impressions	ctr	position
deu	MOBILE	https://www.federweg.com/mtb-karwendelrunde/	karwendelrunde mtb	448	2075	0.21590361445783132	2.8732530120481927
deu	MOBILE	https://www.federweg.com/fox-speedframe-pro-test/	fox speedframe pro test	300	1058	0.2835538752362949	2.0765595463137996
deu	MOBILE	https://www.federweg.com/garmin-fenix-6/garmin-fenix-6s-pro-test/	garmin fenix 6s pro test	293	822	0.3564476885644769	1.1557177615571776
aut	MOBILE	https://www.federweg.com/mtb-karwendelrunde/	karwendelrunde mtb	222	413	0.5375302663438256	1.3898305084745763
deu	MOBILE	https://www.federweg.com/kettenwachs-von-optimize-im-test/	kettenwachs	218	2117	0.10297590930562116	3.6381672177609827
deu	MOBILE	https://www.federweg.com/kettenwachs-von-optimize-im-test/	optimize kettenwachs	208	2403	0.08655846858094049	2.3741156887224304
deu	MOBILE	https://www.federweg.com/garmisch-partenkirchen-enningalm/	enningalm mtb	205	528	0.38825757575757575	2.1382575757575757
deu	MOBILE	https://www.federweg.com/mtb-tour-zur-reintalangerhütte/	reintalangerhütte mtb	196	456	0.4298245614035088	1.0
deu	DESKTOP	https://www.federweg.com/mtb-karwendelrunde/	karwendelrunde mtb	195	1039	0.18768046198267566	2.982675649663138
deu	DESKTOP	https://www.federweg.com/fox-speedframe-pro-test/	fox speedframe pro test	194	649	0.29892141756548535	2.4098613251155623
deu	MOBILE	https://www.federweg.com/fox-speedframe-pro-test/	fox speedframe pro	193	16498	0.011698387683355559	5.817311189235059
deu	MOBILE	https://www.federweg.com/dmr-vault-test/	dmr vault	179	3143	0.05695195672923958	2.582882596245625
deu	MOBILE	https://www.federweg.com/fox-speedframe-pro-test/	fox speedframe mips test	170	446	0.3811659192825112	1.141255605381166

Abb. 1: Exemplarischer Aufbau einer CSV-Datei mit Search-Console-Daten

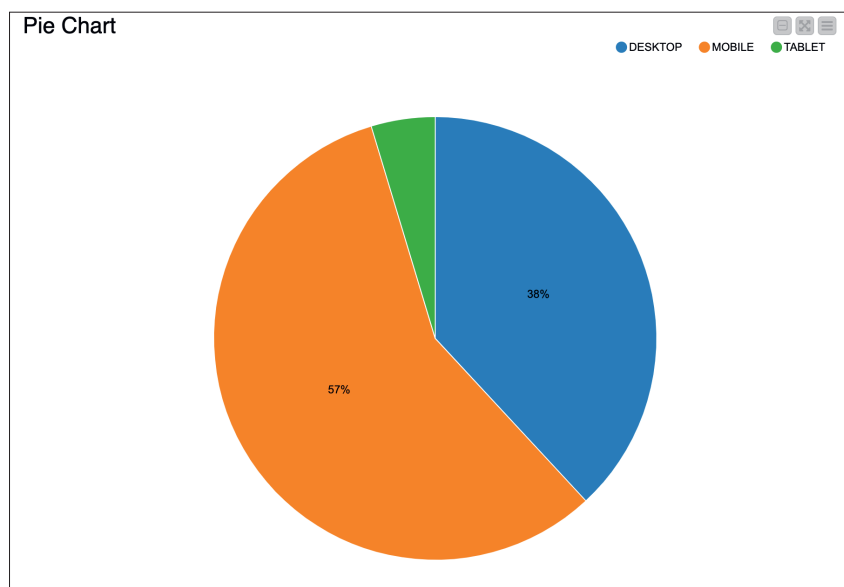


Abb. 2: Prozentuale Verteilung der Klicks auf Gerätekategorien

eigenen Anwendung (zum Beispiel in Python oder R) notwendig.

- » Verwendung von Drittanbieter-Tools (zum Beispiel Supermetrics): Sie sind einfach zu nutzen, verursachen aber zusätzliche Kosten.
- » Bulk-Export von der Search Console in BigQuery: Dabei werden sehr viele Daten bereitgestellt, die sich direkt in BigQuery abfragen oder als CSV exportieren lassen. Allerdings stehen die Daten erst ab dem Tag der Einrichtung zur Verfügung.
- » Looker Studio (ehemals Google Data Studio): In Looker Studio lassen sich die Search-Console-Daten recht einfach importieren und es ist möglich, bis zu 750.000 Zeilen als CSV-Datei zu exportieren.

Es ist zweitrangig, aus welcher Quelle die Daten für die folgende Ana-

lyse stammen. Auch die Anzahl der Spalten in der Tabelle ist nicht entscheidend. Es gilt zwar die Regel „Je mehr, desto besser“, aber auch aus kleineren Tabellen lassen sich interessante Erkenntnisse gewinnen.

Die folgenden Spalten sollten jedoch (mindestens) in Ihren Rohdaten enthalten sein:

- » query: Suchanfragen, die zu einer Website geführt haben. Diese Metrik zeigt, nach welchen Keywords Nutzer suchen, bevor sie auf eine bestimmte Seite gelangen.
- » page: die URL, auf die die Nutzer von den Suchergebnissen aus geklickt haben
- » clicks: die Anzahl der Klicks, die auf eine Website von den Suchergebnissen aus erfolgt sind
- » impressions: die Anzahl der Ein-

blendungen einer Website in den Suchergebnissen für eine bestimmte Suchanfrage

- » position: die durchschnittliche Position einer Website in den Suchergebnissen für eine bestimmte Suchanfrage
- » ctr: Die Klickrate (Click-Through-Rate) gibt an, wie oft Nutzer auf eine Website klicken, nachdem sie sie in den Suchergebnissen gesehen haben. Die in diesem Beispiel verwendete Datei enthält zusätzlich die Spalten:
- » country: das Land, in dem die Suchanfrage gestellt wurde, für die die Metriken erfasst wurden
- » device: das Gerät, von dem aus die Suchanfrage gestellt wurde

Beachten Sie dabei die korrekte Benennung der Spalten wie oben aufgeführt. Die Tabelle sollte Ihnen als CSV-Datei vorliegen.

Search-Console-Daten automatisiert auswerten mit KNIME

Die vorliegende CSV-Datei wird nun in KNIME importiert. Öffnen Sie dazu KNIME und legen Sie einen neuen Workflow an. Am einfachsten importieren Sie die CSV-Datei per Drag-and-drop auf die Arbeitsfläche. KNIME erzeugt dann eine Node vom Typ CSV-Reader auf der Arbeitsfläche, die die Struktur der Datei automatisch erkennt. Schließen Sie das Konfigurationsmenü und führen Sie die Node aus (Rechtsklick auf die Node und dann auf Ausführen klicken).

Die Daten werden nun in KNIME importiert. Für eine erste schnelle Auswertung muss die Tabelle die Spalten country und device enthalten, um die folgenden Informationen zu erhalten:

- » Wie verteilt sich der Traffic prozentual auf die Länder der Suchenden?
- » Wie verteilt sich der Traffic prozentual auf die Endgeräte der Suchenden?

Stehen Ihnen diese Daten nicht zur Verfügung, überspringen Sie diesen Schritt einfach.

Suchen Sie nun im Node Repository nach der Node Pie Chart, ziehen Sie diese auf die Arbeitsfläche und verbinden Sie sie mit dem CSV-Reader an den schwarzen Dreiecken. Klicken Sie doppelt auf die Pie-Chart-Node, um sie zu konfigurieren.

Wählen Sie im Reiter Options „country“ aus, um die Länderauswertung durchzuführen. Bei Aggregation Method wählen Sie „Sum“ und bei Frequency Column „clicks“, um die Summe der Klicks, die über den jeweiligen Gerätetyp kommen, zu ermitteln. Im Tab General Plot Options wählen Sie im Feld Label Type „Percent“. Um das Diagramm zu erstellen, schließen Sie die Konfiguration, klicken Sie mit rechts auf die Node und wählen Sie Execute and Open Views. Es öffnet sich ein neues Fenster mit einem Diagramm (siehe Abbildung 2).

Suchanfragen nach Traffic clustern

In diesem Abschnitt geht es darum, zu erkennen, wie viele High- und wie viele Low-Performer-Kywords es gibt. Dazu wird die Tabelle nach Suchfragen gruppiert, die Summe der Klicks wird ermittelt und Cluster werden gebildet, die zeigen, wie viele Keywords überhaupt zum Traffic beitragen.

Dazu wird eine GroupBy-Node mit dem CSV-Reader verbunden und wie folgt konfiguriert: Im Tab Groups wird die Spalte query hinzugefügt, während im Tab Manual Aggregation konfiguriert wird (siehe Abbildung 3).

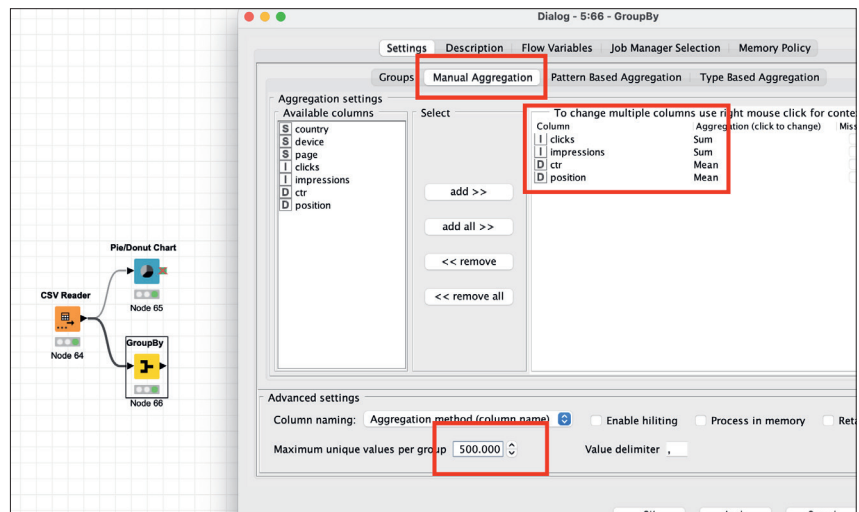


Abb. 3: Konfiguration der GroupBy-Node

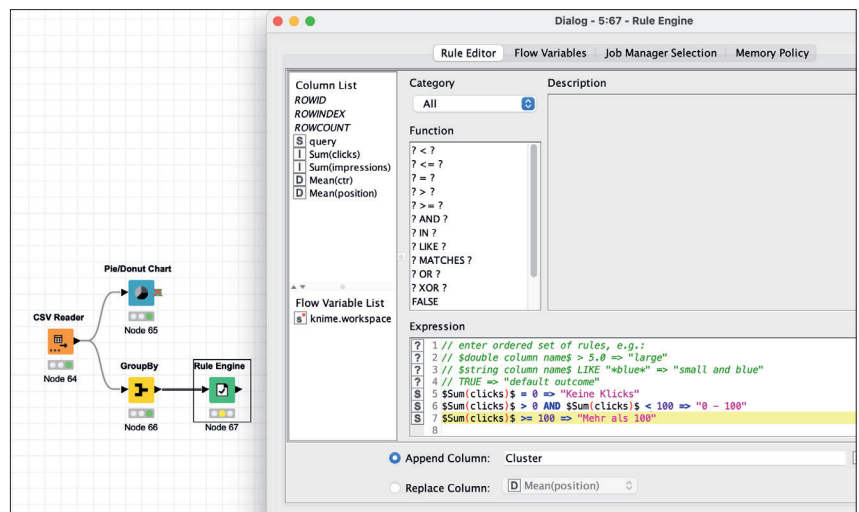


Abb. 4: Erstellen einer Clustering-Regel

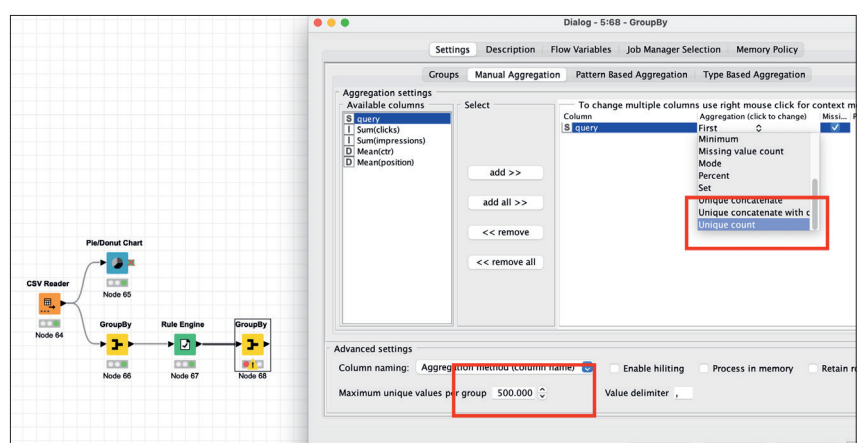


Abb. 5: Gruppierung nach Cluster

Um die Cluster zu bilden, wird mit der Node Rule Engine eine Regel erstellt, die die Keywords nach Klicks gruppiert. Doch keine Angst, Sie müssen dabei nicht mit komplizierten Schleifen arbeiten, es reicht, eine einfache Wenn-dann-

Regel zu erstellen. Das Ergebnis dieser Regel wird anschließend in eine neue Spalte geschrieben.

Zunächst verbinden Sie die Rule Engine mit der GroupBy-Node und öffnen die Konfiguration. Um es einfach zu

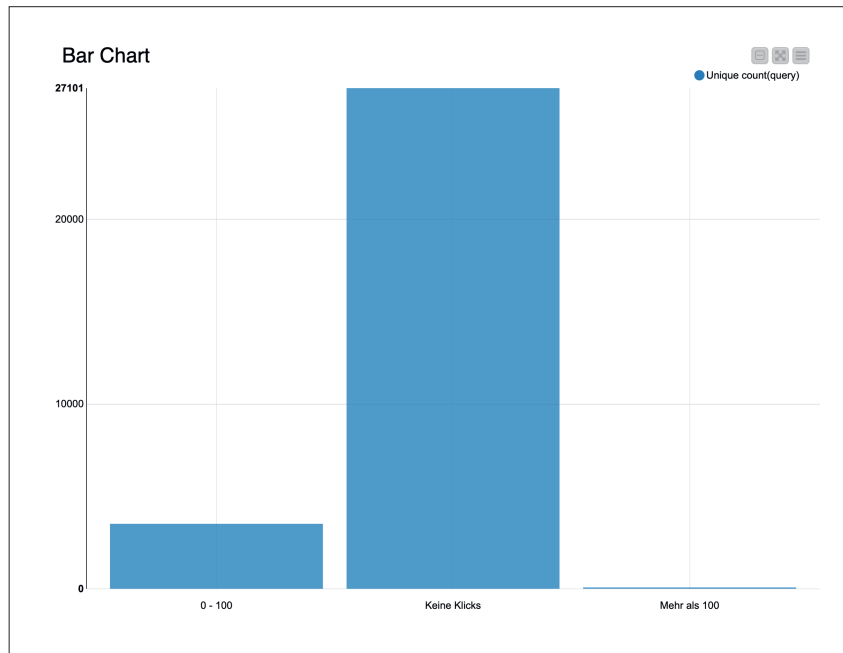


Abb. 6: Die Keywords gruppiert nach Leistung

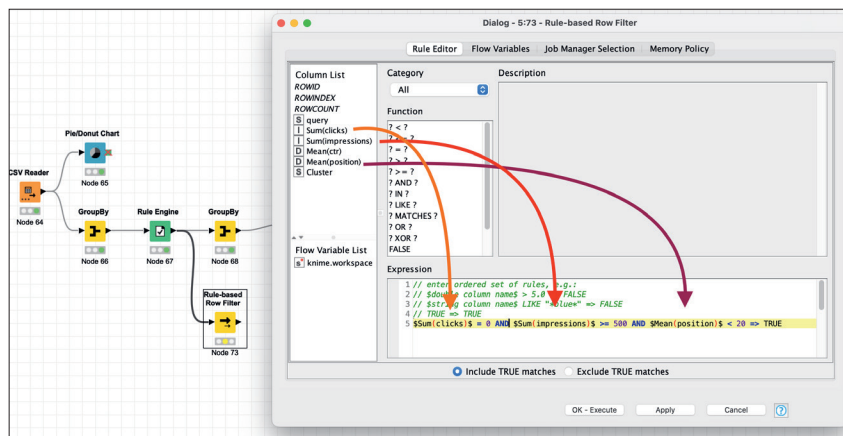


Abb. 7: Konfiguration des Row Filters

halten, reichen zunächst drei Cluster:

- » Keywords, die keine Klicks gebracht haben
- » Keywords, die zwischen einem und 100 Klicks gebracht haben
- » Keywords, die über 100 Klicks gebracht haben

Je nach Größe der Website kann es sinnvoll sein, hier granularer oder mit mehr Clustern zu arbeiten. Das erste Cluster wird gebildet, indem per Doppelklick die Spalte mit den Klicks Sum(clicks) ausgewählt wird, sodass diese im Feld Expression erscheint. Anschließend kann mit der Funktion = 0 gesagt werden, dass diese für alle Zeilen gelten soll, in der der Wert Sum(clicks) 0 entspricht. Mit dem Zusatz => „Keine Klicks“ wird der Engine gesagt, was in die neue Spalte geschrieben werden soll. Die vollständige Konfiguration sehen Sie in Abbildung 4.

Damit sichtbar wird, wie viele Keywords in jedem Cluster enthalten sind, kommt erneut die Node GroupBy zum Einsatz. Es wird dieses Mal nach der neuen Spalte mit den Clustern gruppiert. Aggregiert wird die Spalte query mit der Methode „unique count“, sie zählt, wie viele Keywords pro Cluster enthalten sind, ohne dabei eventuell doppelt vorhandene mehrfach zu zählen. Bei Seiten mit vielen Keywords ist es wichtig, den

Jetzt einschalten:

https://websiteboosting.tv

Browser Plugins
Datenübertragungen...

111 Aufrufe • vor 2 Monaten

MUM und die Google-Suche
der Zukunft - Website...

425 Aufrufe • vor 3 Monaten

SEO Basics: Ist die Anzahl
Worte ein Rankingfaktor?

158 Aufrufe • vor 5 Monaten

SEO Basics Teil 3: Die
Überschriften Hx (H1-H6)

243 Aufrufe • vor 5 Monaten

Ausgabe 70 R Leuchtungen
Teil 9 Google Titel Update

54 Aufrufe • vor 5 Monaten

Ausgabe
SEO Teil...

109 Aufrufe • vor 5 Monaten

Wert Maximum Unique Values per Group hochzusetzen, sonst kann jedes Cluster maximal 10.000 Keywords enthalten (siehe Abb. 5).

Eine Bar-Chart-Node kann nun dazu verwendet werden, das Ergebnis zu visualisieren. Dabei muss in der Konfiguration lediglich der Wert im Bereich Aggregation Method von „Occurrence Count“ zu „Sum“ gewechselt werden. Im Chart wird jetzt sichtbar, wie viele Keywords zum Traffic der Website beigetragen haben. Vor allem wird aber deutlich, wie viele Keywords nicht zum Traffic beigetragen haben (siehe Abbildung 6).

Ob Potenziale bei den O-Klick-Keywords vorhanden sind, wird im nächsten Schritt analysiert.

Potenzial-Keywords finden

Zunächst muss die Tabelle so gefiltert werden, dass in ihr nur noch O-Klick-Keywords enthalten sind. Dazu wird dem Workflow eine Rule-based-Row Filter-Node hinzugefügt und mit der Rule-Engine-Node verbunden. Mit dieser Node lässt sich mit einfachen Regeln eine Filterlogik erstellen. Deren Ziel ist es, ausschließlich Keywords in der Tabelle zu behalten, die:

- » keine Klicks gebracht haben.
- » mehr als 500 Impressionen hatten.
- » ein besseres Ranking als Position 20 aufweisen.

In den Einstellungen des Rule-based Row Filter wird der Filter wie folgt angelegt: $\$Sum(clicks)\$ = 0 \text{ AND } \$Sum(impressions)\$ \geq 500 \text{ AND } \$Mean(position)\$ < 20 \Rightarrow TRUE$. Wie bei der Rule Engine lassen sich die Spaltennamen einfach per Doppelklick in das Feld mit der Regel einfügen (siehe Abbildung 7).

Zur Visualisierung der Daten eignet sich die Node Scatter Plot des Anbieters Plottly, sie kann einfach per Drag-and-drop von [einfach.st/knime4538](https://www.knime.com/plugins/plotly) installiert werden. Konfiguriert werden kann die Node wie folgt: X-Axis Column: Sum(impressions), Y-Axis Column: Mean(Position), GroupBy Column: query. Um etwas

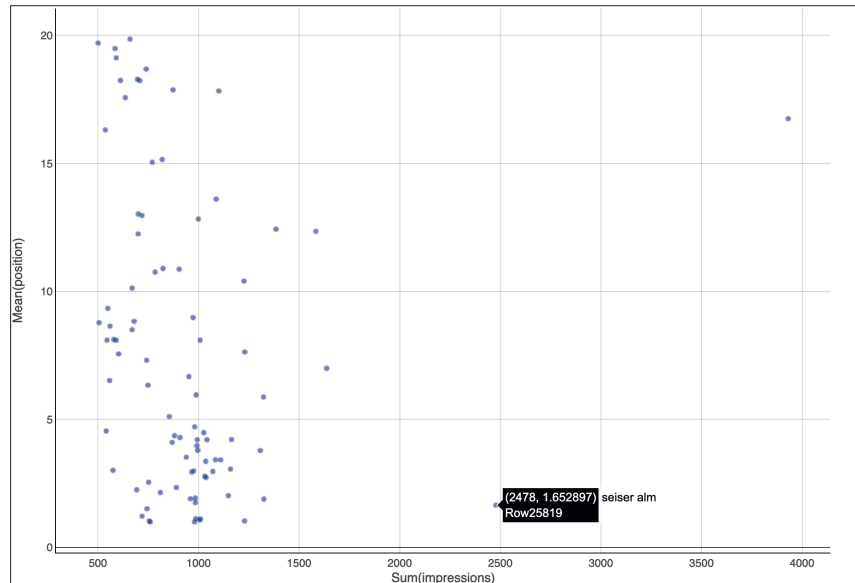


Abb. 8: Die O-Klick-Keywords in einem Scatter Plot

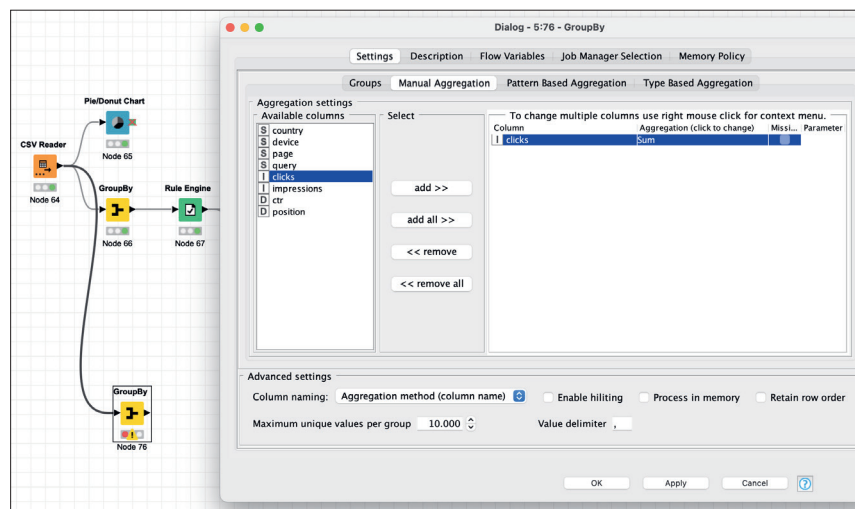


Abb. 9: Nach URL gruppieren und die Summe der Klicks bilden

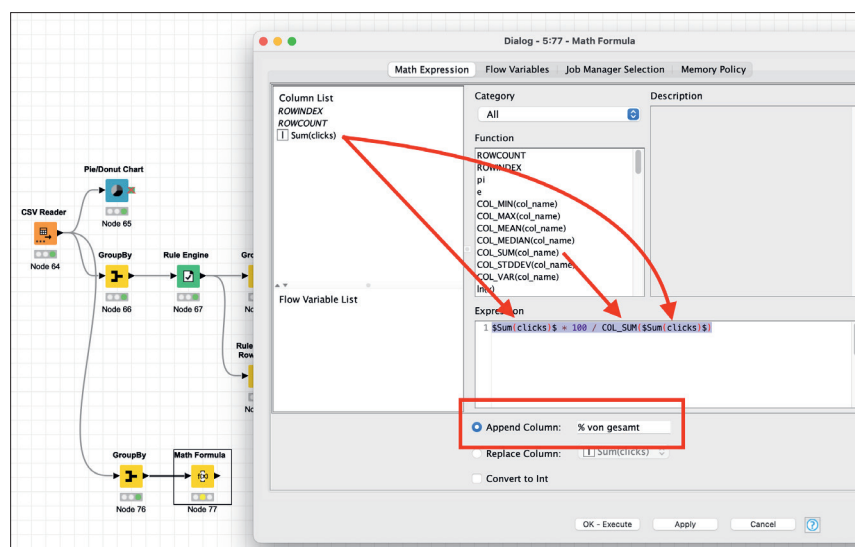


Abb. 10: Ermitteln des Traffic-Anteils jeder URL am Gesamt-Traffic

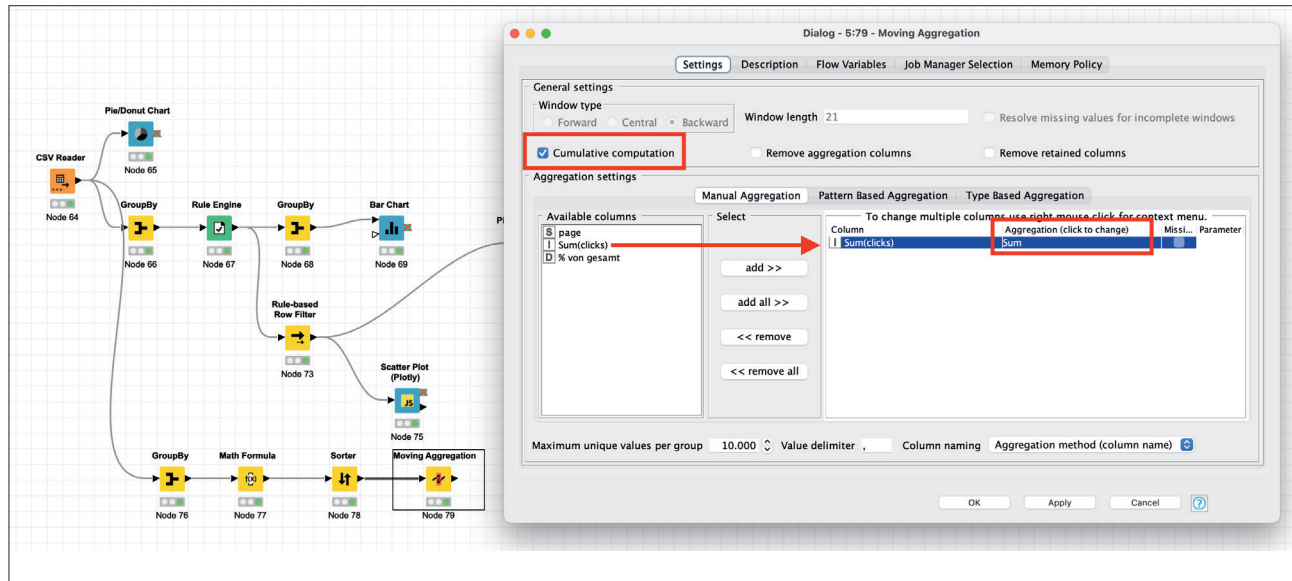


Abb. 11: Konfiguration der Moving Aggregation

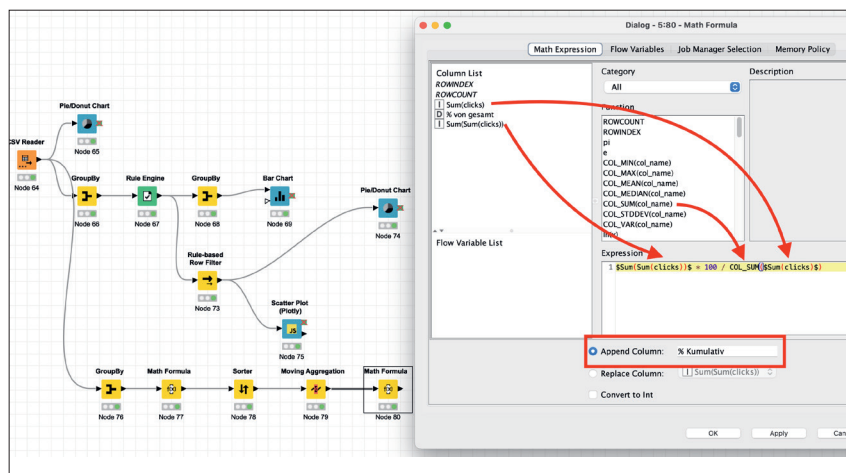


Abb. 12: Berechnen der kumulativen Prozente

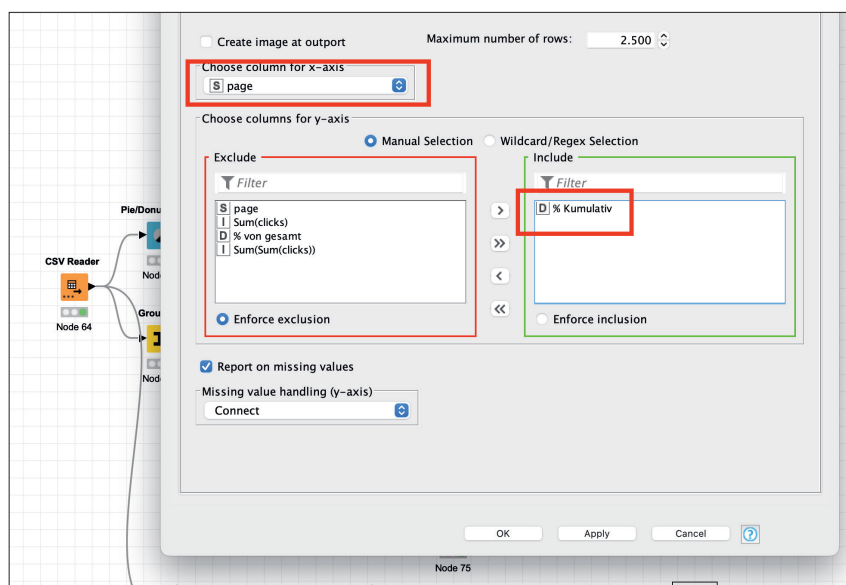


Abb. 13: Konfiguration des Liniendiagramms

Farbe in die Grafik zu bringen, können im Tab General Plot Options Farbe und Größe des Charts verändert werden. Die Grafik zeigt nun jedes Keyword als einen Datenpunkt verteilt nach Impressionen (x-Achse) und Position (y-Achse). Je weiter rechts unten ein Keyword erscheint, desto mehr Potenzial hat es, weil es viele Impressionen und ein gutes Ranking hat und trotzdem (noch) keine Klicks gebracht hat (siehe Abbildung 8).

Den Beitrag einzelner URLs am Gesamt-Traffic ermitteln

Die wichtigsten URLs einer Website sind den zuständigen Personen in der Regel bekannt. Wie hoch ihr Anteil am gesamten Google-Traffic ist, kann eine wichtige Information sein, wenn es beispielsweise bei einem Relaunch um das Thema Weiterleitungen geht. Nach dem Pareto-Prinzip sollten 20 % Redirects reichen, um 80 % des Traffics zu erhalten. Doch stimmt das?

Um diese Frage zu beantworten, müssen Sie zunächst ein paar Schritte zurückgehen. Verbinden Sie eine GroupBy-Node mit dem CSV-Reader und gruppieren Sie die Tabelle nach der Spalte Page. Im Bereich Manual Aggregation der GroupBy-Node wählen Sie die Spalte clicks und als Aggregation

Methode Sum (siehe Abbildung 9). Tipp: Wenn Sie zusätzlich die Spalte query wählen und als Aggregation Methode unique count, erhalten Sie auch die Information, für wie viele Keywords eine URL rankt.

Im nächsten Schritt ermitteln Sie mithilfe der Node Math Formula, zu wie viel Prozent jede URL zum Gesamt-Traffic beiträgt. Dazu wird die Anzahl an Klicks jeder URL mit 100 multipliziert und durch die Gesamtsumme des Traffics geteilt. Glücklicherweise stellt die Node die Funktion COL_SUM zur Verfügung, die diese Gesamtsumme automatisch ermittelt. So lautet die Math Formula lediglich: $\$Sum(clicks)\$ * 100 / COL_SUM(\$Sum(clicks)\$)$. Das Ergebnis wird in eine neue Spalte mit dem Titel „% von gesamt“ geschrieben (siehe Abbildung 10).

Die neue Spalte zeigt nun, wie groß der Anteil des Traffics wäre, der verloren gehen würde, wenn eine dieser URLs keine Klicks mehr bekäme. Mit der Node Sorter lässt sich die neue Spalte absteigend sortieren, sodass die wichtigsten URLs sofort sichtbar sind.

Diese Sortierung ist auch für den nächsten Schritt wichtig. Mit der Node Moving Aggregation kann eine zusätzliche, kumulative Spalte erzeugt werden, die die Klicks der einzelnen URLs Zeile für Zeile addiert, bis in der untersten Zeile die Gesamtsumme aller Klicks steht. Dies ist eine Hilfsspalte für weitere Berechnungen. Wählen Sie dazu in der Konfiguration der Node bei den General Settings „Cumulative computation“, bewegen Sie die Spalte Sum(Clicks) in das rechte Feld und wählen Sie als Aggregation Methode Sum (siehe Abbildung 11).

Mithilfe dieser neuen Spalte lässt sich nun mit einer weiteren Math Formula eine zusätzliche Spalte erzeugen, die das Gleiche in Form prozentualer Werte durchführt. Dazu wird die neue Spalte mit dem sperrigen Namen Sum(Sum(clicks)) mit 100 multipliziert

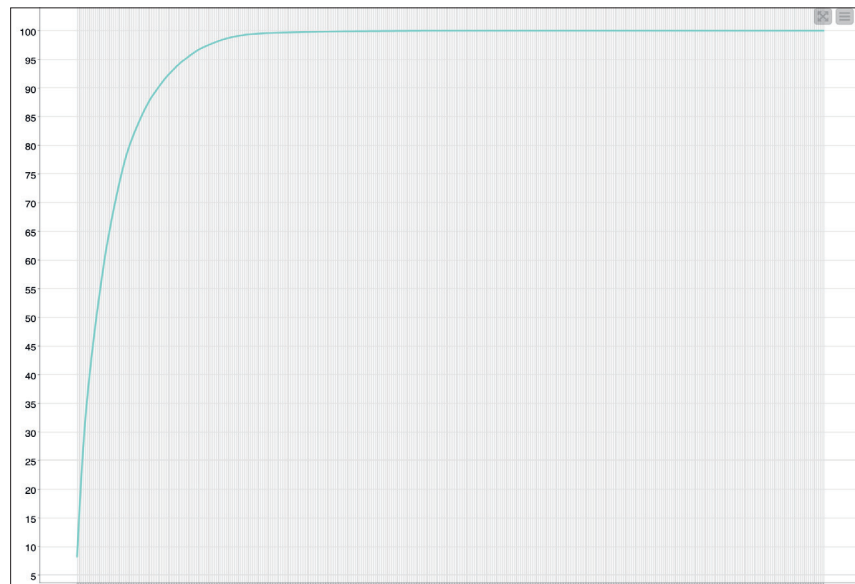


Abb. 14: Das fertige Pareto-Chart

und durch die Summe des Gesamt-Traffics geteilt: $\$Sum(Sum(clicks))\$ * 100 / COL_SUM(\$Sum(clicks)\$)$. Das Ergebnis wird in eine weitere Spalte geschrieben, beispielsweise mit dem Titel % Kumulativ (siehe Abbildung 12).

Dank dieser Informationen lässt sich nun die Frage beantworten, wie viele Klicks verloren gehen würden, wenn tatsächlich nur 20 % aller Seiten weitergeleitet würden. Umgekehrt lässt sich aber auch die Frage beantworten, wie viele Weiterleitungen wir denn brauchen, wenn wir 80 % des Traffics behalten möchten.

Diese Tabelle lässt sich auch sehr anschaulich in einem Pareto-Chart visualisieren. Schließen Sie dazu einfach eine Node mit dem Namen Line Plot an die Math Formula an. Konfigurieren Sie sie so, dass die x-Achse die URLs (Spalte „page“) enthält und auf der y-Achse die Werte der Spalte „% kumulativ“ abgebildet werden (siehe Abbildung 13).

Im Endergebnis sehen Sie nun ein Liniendiagramm, in dem die Klickzahlen einer jeden URL kumuliert werden. Damit haben Sie einen schnellen Überblick darüber, ob die Website ihren Traffic gleichmäßig über viele URLs bekommt oder von einigen starken Seiten überproportional profitiert. Denn: Je steiler das Diagramm, desto ungleicher verteilt sich der Traffic auf die Gesamt-

zahl der URLs und je länger die Linie auf Höhe der 100%-Marke ist, desto mehr Ballast befindet sich auf der Website, der nicht zum Traffic beiträgt (siehe Abbildung 14).

Der Vollständigkeit halber sei erwähnt, dass in diesem Beispiel nur die URLs in die Auswertung einfließen, die der Search Console bekannt sind. Für ein vollständigeres Bild (das in aller Regel weniger schmeichelhaft ausfällt), sollte die Pareto-Auswertung auf Basis aller (indexierbaren) URLs der Website durchgeführt werden.

Fazit

Auch wenn das Erstellen dieses Workflows etwas Zeit in Anspruch nimmt, haben Sie anschließend ein Werkzeug in der Hand, das Ihnen schnell Informationen über Search-Console-Daten liefert, die Ihnen das Webinterface der GSC nicht bietet. Mit den Visualisierungen sind Sie in der Lage, die Ergebnisse schnell an Kunden oder Vorgesetzte zu kommunizieren. Das Schöne dabei: Dieser Workflow ist zu 100 % auf jede Website anwendbar, zu der Sie einen Search-Console-Zugriff haben.

Den fertigen Workflow können Sie kostenlos unter einfach.st/knime79 herunterladen. ¶