

Thomas Kaiser

The oil of Qlik – Big SEO visualisieren

Ein Bild sagt mehr als 1.000 Worte oder 100.000 Daten. Doch Diagramme zu erstellen, die nicht nur eine klare Aussage abbilden, sondern aus denen man auch belegbare Schlussfolgerungen ziehen kann, ist schwierig. Thomas Kaiser vermittelt Ihnen die Grundlagen für das richtige Vorgehen und zeigt mittels der Software Qlik an Beispielen die vielfältigen Möglichkeiten auf.



DER AUTOR



Thomas Kaiser ist Gründer und Geschäftsführer der cyberpromote GmbH und der Forecheck LLC. Er ist passionierter Entwickler, Musiker und SEO-Gourmet, schreibt Bücher und Fachartikel

Big Data kommt langsam auch in der SEO-Szene an und Datenanalysten entdecken die SEO-Daten für Auswertungen im eigenen Unternehmen. Hierfür gibt es verschiedene Tools, das Zusammenführen der Daten ist aber heute oft nur durch manuelles Übertragen und Anpassen möglich. Die Kombination mit weiteren Unternehmensdaten ermöglicht noch tiefere Einblicke.

Big Data ist primär im Bereich von Unternehmenszahlen und im CRM verankert, und damit soll es möglich sein, in die Zukunft zu

blicken, Stichwort „Predictive Analytics“. Im SEO- und SEA-Bereich ist man ja schon glücklich, wenn man die Vergangenheit versteht und daraus Schlüsse für mehr Erfolg ziehen kann.

Heute gibt es leistungsfähige Tools, die in der Lage sind, große Mengen an Daten aus verschiedenen Quellen zu verarbeiten, zu verknüpfen, aufzubereiten und grafisch auszuwerten. Hierzu gehören Tools wie Rapidminer (www.rapidminer.com), Tableau (www.tableau.com), Spotfire (spotfire.tibco.com), Pentaho (www.pentaho.com) oder Talend (www.talend.com).

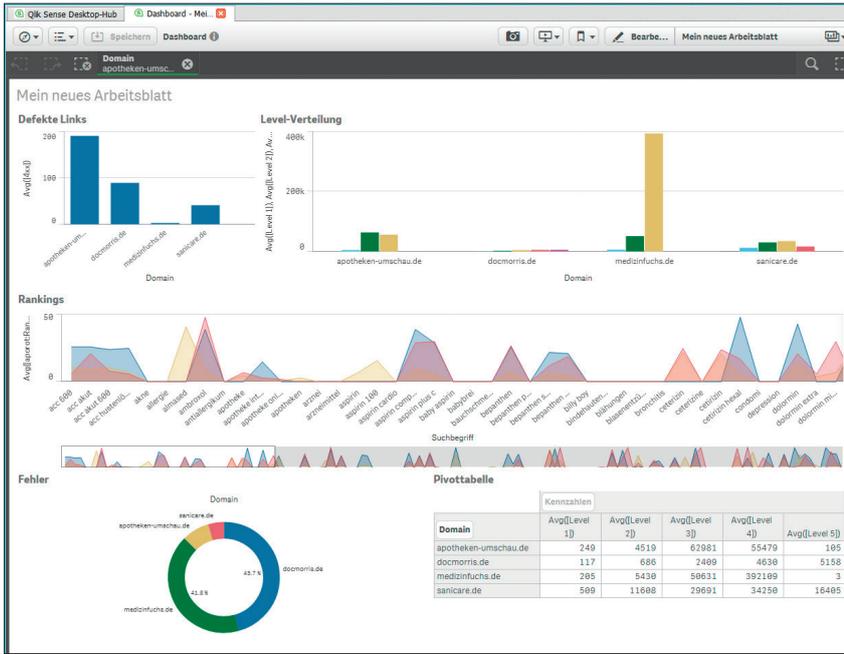


Abb. 1: Dashboard mit einigen SEO-Kennzahlen

TIPP
 Unter www.forecheck.de/download/wbqlik/ können alle Dateien für diese Beispiele hier heruntergeladen werden. Zudem gibt es dort noch ein paar weitere Daten, mit denen man für den Anfang etwas experimentieren kann.

Qlik ist ebenfalls ein Anbieter mit langer Historie, der mit Qlik Sense Desktop eine einfache Windows-Software mit intuitiver Oberfläche anbietet, die zudem kostenlos ist und nach der englischen Installation eine deutsche Oberfläche besitzt. Sie müssen nur ein paar Daten über sich preisgeben und können die Software dann herunterladen (<http://einfach.st/qlik>). Doch zunächst sollen ein paar Grundlagen betrachtet werden, bevor an Beispielen die Möglichkeiten von Qlik aufgezeigt werden.

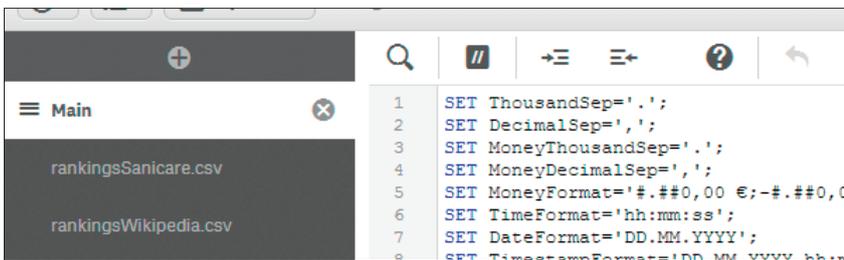


Abb. 2: Die Grundeinstellungen in Qlik

Korrelation und Kausalität

Die SEO-Szene setzte sich oft über wissenschaftliche Grundlagen hinweg. Spektakulär war sicherlich die Behauptung, Facebook Shares wären der wichtigste Rankingfaktor. Immerhin entschuldigte sich Rand Fishkin öffentlich dafür, dass man auch bei SeoMoz fälschlicherweise von Rankingfaktoren gesprochen hatte (<http://einfach.st/rand5>).

Eine Korrelation sagt nichts über Kausalität, sprich Ursache und Wirkung aus. Wer also Daten auswertet und vergleicht und eine Behauptung aufstellt, sollte das Beweise liefern. Doch selbst nachvollziehbare Beweise müssen eine Kausalität nicht zwingend belegen. Als Groupon in einem spektakulären Experiment feststellte, dass etwa 60 % des Traffics über den direkten Kanal eigentlich organisch sind, waren die Belege nachvollziehbar (<http://einfach.st/sel6>), die Argumente und Fragen in den Kommentaren allerdings auch.

Es gibt oft Einflussfaktoren, die man nicht im Blick hat oder nicht kontrollieren kann. Diese können die Datenanalyse unbrauchbar machen oder sogar zu falschen Schluss-

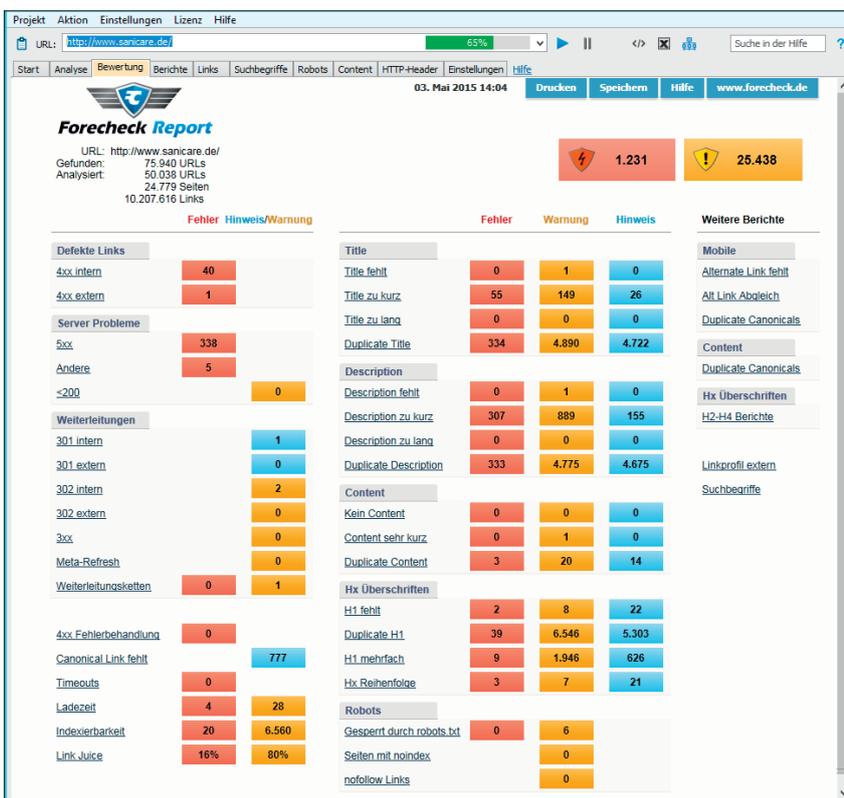


Abb. 3: Forecheck-Bewertung für sanicare.de

folgerungen führen. Erfahrung und gesunder Menschenverstand sind sehr wichtig.

Grundlagen

Für eine Auswertung von Daten muss man zunächst verstehen, welche Daten man hat, wie man diese auswertet und welche Fragestellung man beantworten möchte. Zudem sollte man natürlich wissen, welche Einflussfaktoren für die Analyse relevant sind, was gerade bei SEO schwierig ist. Möchte man beispielsweise Suchbegriffe analysieren, sollte man nicht diejenigen auswerten, mit denen man gut platziert ist. Man erhält nur dann aussagekräftige Daten, wenn man die Begriffe mit einschließt, zu denen man nicht gefunden wird. Hierfür sollte man eine möglichst vollständige Liste erstellen, die für die eigene Website alle relevanten Suchbegriffe beinhaltet.

Probleme können die Formate der Daten bereiten. Gerade die Unterschiede zwischen deutschen und englischen Zahlen- und Datumsformaten sind tückisch. Grundsätzlich versteht Qlik diese Formate automatisch. Wenn Sie aber verschiedene Formate aus einer Datenquelle importieren, können Fehler auftreten. Beachten Sie auch, dass manche Zahlen als Text formatiert sind. Dateigrößen können als 3.467 Byte oder 3,39 kB oder 0,0033 MB angegeben sein. Dann muss man nicht nur die Maßeinheit aus dem Feld entfernen, sondern eventuell auch noch umrechnen. Grundsätzlich bietet Qlik auch dabei umfassende Unterstützung. Die Grundeinstellungen für Formate finden Sie im Dateneditor, erreichbar über den Button links oben.

Click mich

Qlik Sense Desktop ist noch ein recht junges Produkt des Unternehmens Qlik (www.qlik.com), welches schon seit 1993 im Markt für Daten-



Abb. 4: Der Qlik-Datenimport



Abb. 5: Eingebettete Feldnamen

visualisierung tätig ist. Das Ziel war ein einfacher Einstieg mit intuitiver Bedienung und mächtigen Funktionen. Man kann in Qlik eine App anlegen und darin Arbeitsblätter erstellen. In diesen kann man die Daten in vielen verschiedenen Diagrammen ausgeben. Die Apps lassen sich in der Cloud für einen öffentlichen Zugriff freigeben. Jede erstellte Grafik kann im sogenannten Hub als Mastervorlage gespeichert werden und steht damit

in jeder anderen App zur Nutzung zur Verfügung.

Grundsätzliche kann Qlik Daten aus Excel-, CSV-, HTML- und XML-Dateien lesen, sich aber auch mit verschiedensten Datenbanken und Webdiensten verbinden. Für ein erstes Beispiel wollen wir sanicare.de genauer betrachten. Die Excel-Datei Bewertung_Vergleich.xls beinhaltet wichtige Kennzahlen von sanicare.de und drei weiteren Domains, die im Umfeld von Online-Apotheken auftauchen: *apotheken-umschau.de*, *docmorriss.de* und *medizinfuchs.de*. Hierzu wurden bei allen vier Domains die ersten 50.000 URLs analysiert; dies soll für eine Auswertung genügen. Ein Blick

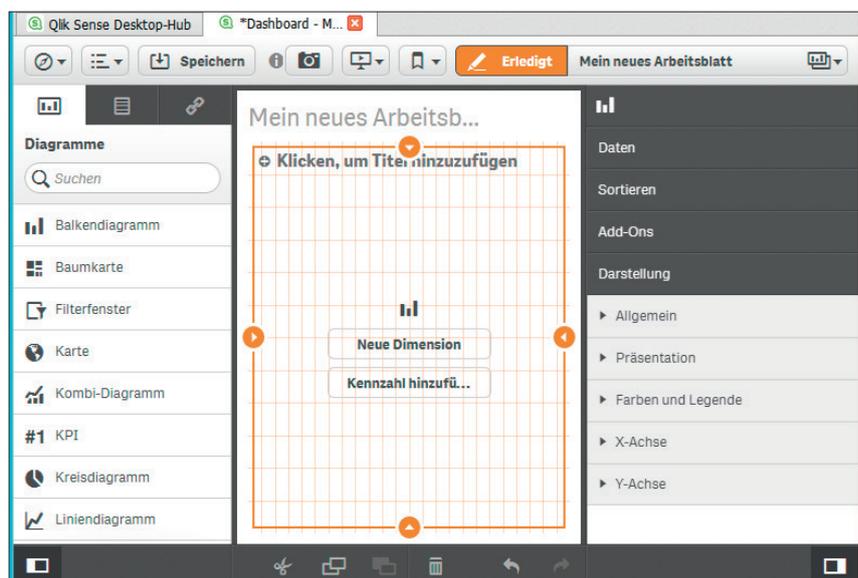


Abb. 6: Das erste Diagramm in Qlik



Abb. 7: Vorhandene Felder

auf die Analyse von sanicare.de zeigt, dass es hier typische SEO-Probleme gibt. Im Dokument Bewertungen.doc können Sie die Analysen der anderen Domains ebenfalls ansehen.

Für eine erste App starten Sie Qlik, erstellen eine neue App und dann ein neues Arbeitsblatt. Beim Start wird sich vermutlich die Firewall melden, damit Sie für Qlik einen Port öffnen (dies ist nicht zwingend notwendig, Sie können aber Apps in die Cloud kopieren und mit anderen zusammen daran arbeiten). Ziehen Sie die Datei Bewertung_Vergleich.xls einfach auf das Arbeitsblatt, dann öffnet sich der Datenimport. Diese Datei beinhaltet wichtige Kennzahlen aus Forecheck für alle vier Domains. Sie können die Spalten einzeln abwählen und durch einen Klick auf den Spaltennamen diesen bei Bedarf auch ändern. Für unsere Analyse müssen Sie hier nichts anpassen.

Wenn die Daten Überschriften für die Spalten haben, Qlik diese aber

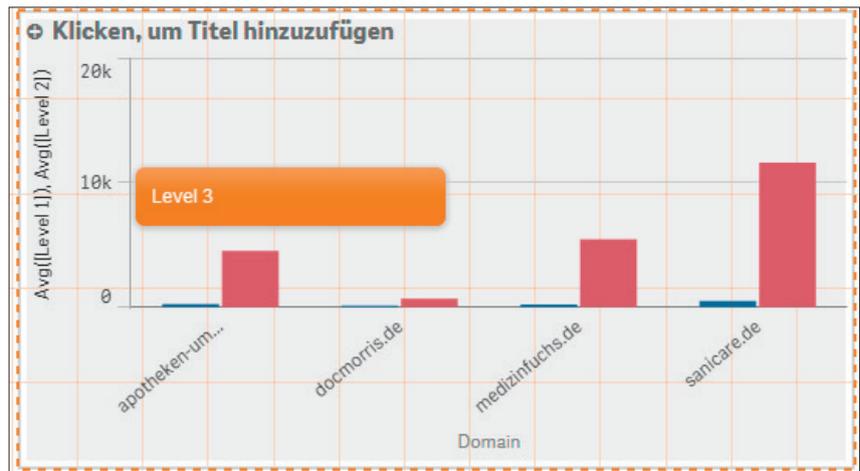


Abb. 8: Kennzahl per Drag & Drop hinzufügen

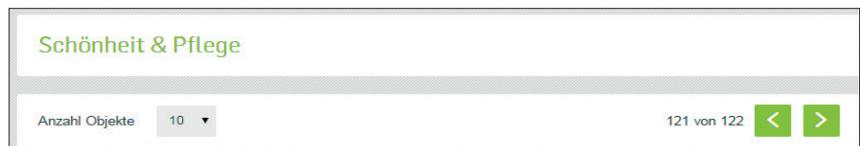


Abb. 9: Paginierung bei docmorriss verschonkt viel SEO-Potenzial

nicht erkennt, müssen Sie bei der Option „Feldnamen“ auf „eingebettete Feldnamen“ umstellen.

Das erste Diagramm

Klicken Sie auf „Daten laden“ und anschließend auf „Arbeitsblatt bearbeiten“. Klicken Sie jetzt links auf „Balkendiagramm“ und ziehen Sie es auf das Arbeitsblatt. Nun müssen noch eine Dimension und eine Kennzahl hinzugefügt werden.

Eine Dimension steht für die Werte auf der X-Achse, die stets aggregiert verarbeitet werden. Wenn Sie also große Datenmengen verarbeiten, die sinnvoll visualisiert werden sollen, müssen die Werte zusammengefasst werden. Hier gibt es ein paar Standardfunktionen wie AVG (Average = Mittelwert) oder SUM (Summe). Später folgt noch ein sinnvolles Beispiel hierzu. Zunächst klicken Sie auf „Neue Dimension“ und wählen aus der Liste „Domain“ aus. Als Kennzahl fügen Sie „4xx“ hinzu. Sie sehen nun in der Grafik den Vergleich der vier Domains hinsichtlich der defekten Links (= Status 4xx; siehe Abbildung 1).

Die erste Grafik ist geschafft. Nun fügen wir noch weitere hinzu, bis wir ein Arbeitsblatt haben, wie es Abbildung 1 zeigt. Im Balkendiagramm oben rechts (Level-Verteilung) sind mehrere Kennzahlen eingefügt. Um Kennzahlen hinzuzufügen, ziehen Sie diese einfach aus dem Tab „Felder“ links auf die Grafik.

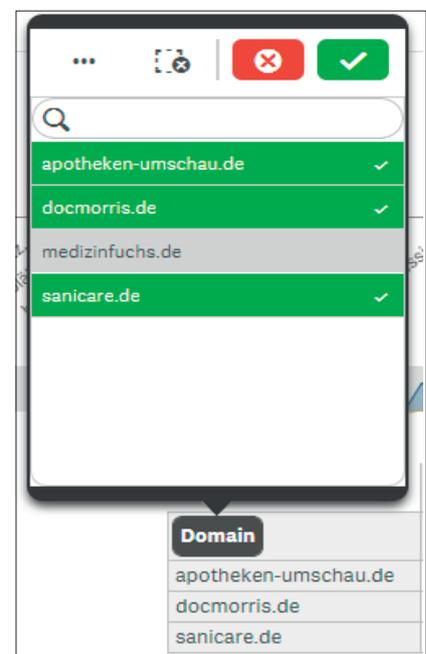


Abb. 10: Die dynamische Pivottable

Wenn Sie die Kennzahl auf die Grafik ziehen, können Sie diese der Grafik hinzufügen.

Als Aggregation wählen Sie die Funktion AVG (Mittelwert). Fügen Sie nach und nach die Felder Level 2 bis Level 5 hinzu. Sieht man sich die fertige Grafik an, erkennt man, dass docmorris die ungünstigste Verteilung hat. In Forecheck kann man in der Analyse sogar Seiten bis zu einem Level von 123 finden. Das bedeutet, dass der schnellste Weg zu dieser Seite über 123 Klicks von der Startseite aus führt. Der Grund lässt sich schnell erkennen. Die Paginierung ist so ungünstig gewählt, dass man Produkte nur durch endloses Klicken von Seite zu Seite erreicht. Deswegen liegen viele Seiten von docmorris auf einem Level jenseits von 5 und können in dem Diagramm gar nicht mehr abgebildet werden.

Für das Liniendiagramm in der Mitte von Abbildung 1 müssen Sie noch die Dateien Rankings_sanicare.csv, Rankings_aporot.csv und Rankings_docmorris.csv importieren. Als Dimension für das Liniendiagramm wählen Sie „Suchbegriff“, als Kennzahlen wählen Sie die Felder „aporot Rankings“, „sanicare Rankings“ und „docmorris Rankings“. In der Grafik fällt auf, dass sich die rote und die blaue Linie stark ähneln, die gelbe eher abweicht.

Natürlich macht es keinen Sinn, dass eine schlechte Platzierung einen starken Ausschlag in der Grafik ergibt wie hier. Sie können die Daten natürlich anpassen. Qlik bietet sehr viele Funktionen für den Import der Daten im Dateneditor (erreichbar über den Button oben links). Beachten Sie das Konzept von Qlik: Felder mit gleichem Namen werden beim Import automatisch verknüpft. Wenn es nun aber pro Datensatz mehr als einen gleichen Spaltennamen gibt, kann das Probleme bereiten. Dann warnt Qlik

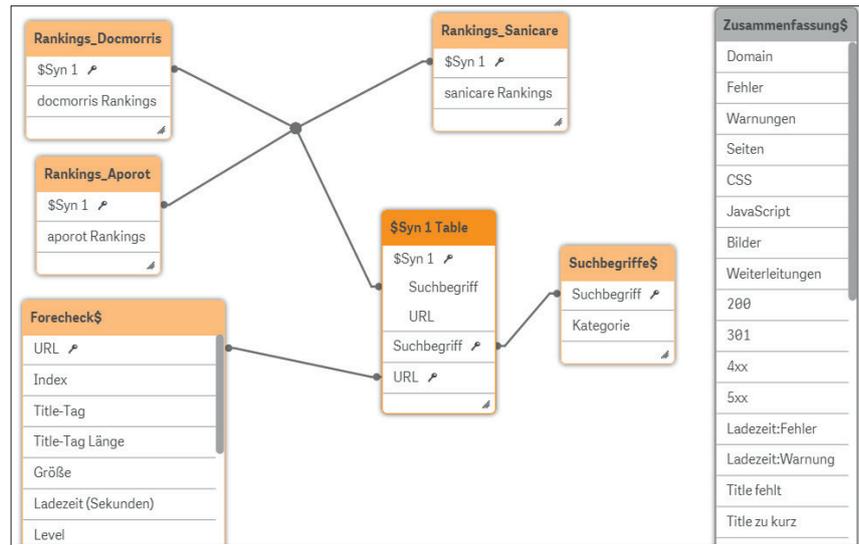


Abb. 11: Das Datenmodell zeigt, wo Daten verknüpft sind

mit dem Einsatz von „synthetischen Schlüssel“.

Daher haben die Spalten der Rankings in den Dateien zu den Domains nicht den gleichen Namen, sondern beinhalten den Domainnamen, z. B. „sanicare Rankings“. Damit werden die Platzierungen nicht automatisch verknüpft. Im SEO-Bereich sind die Suchbegriffe und die URL typische Verknüpfungen zwischen Daten.

Magische Pivottabellen

Die Pivottabelle unten rechts in der Grafik aus Abbildung 1 besteht aus der Dimension „Domain“ und den Kennzahlen „Level 1“ bis „Level 5“. Wenn Sie den Editiermodus verlassen und in der Pivottabelle auf den Button „Domain“ klicken, können Sie die einzelnen Domains entfernen; alle Grafiken im Arbeitsblatt passen sich dann automatisch der Auswahl an!

Für einen Einblick in den Markt der Online-Apotheken finden Sie im Dokument Mitbewerbervergleich.xls eine Liste der Domains, die für 150 ausgewählte Begriffe in den Suchergebnissen auftauchen, gelistet in der Reihenfolge der absteigenden Sichtbarkeit in Google. Sie zeigt, dass vor allem Informationsportale wie apotheken-umschau.de und onmeda.de bei den



Abb. 12: Spaltenzahl einstellen

Begriffen ganz vorne sind. Die Begriffe finden Sie in der Datei Suchbegriffe.xls; sie sind segmentiert in die drei Gruppen „Produkt“ für produktbezogene Begriffe („Nasivin“, „Schwangerschaftstest“), problemorientierte Begriffe („Pickel“, „Durchfall“) und generische Suche („Naturkosmetik“, „Apotheke Online“). Natürlich ist die Liste nicht vollständig, sie soll für die Beispiele hier aber genügen.

Sie zeigt, dass vor allem Informationsportale wie apotheken-umschau.de und onmeda.de bei den Begriffen ganz vorne sind.

Wenn Sie die Datei Suchbegriffe.csv dem Arbeitsblatt hinzufügen, wird diese über die Spalte „Suchbegriffe“ mit den anderen Daten automatisch verknüpft. Hierzu wird ein synthetischer Schlüssel erzeugt, da diese

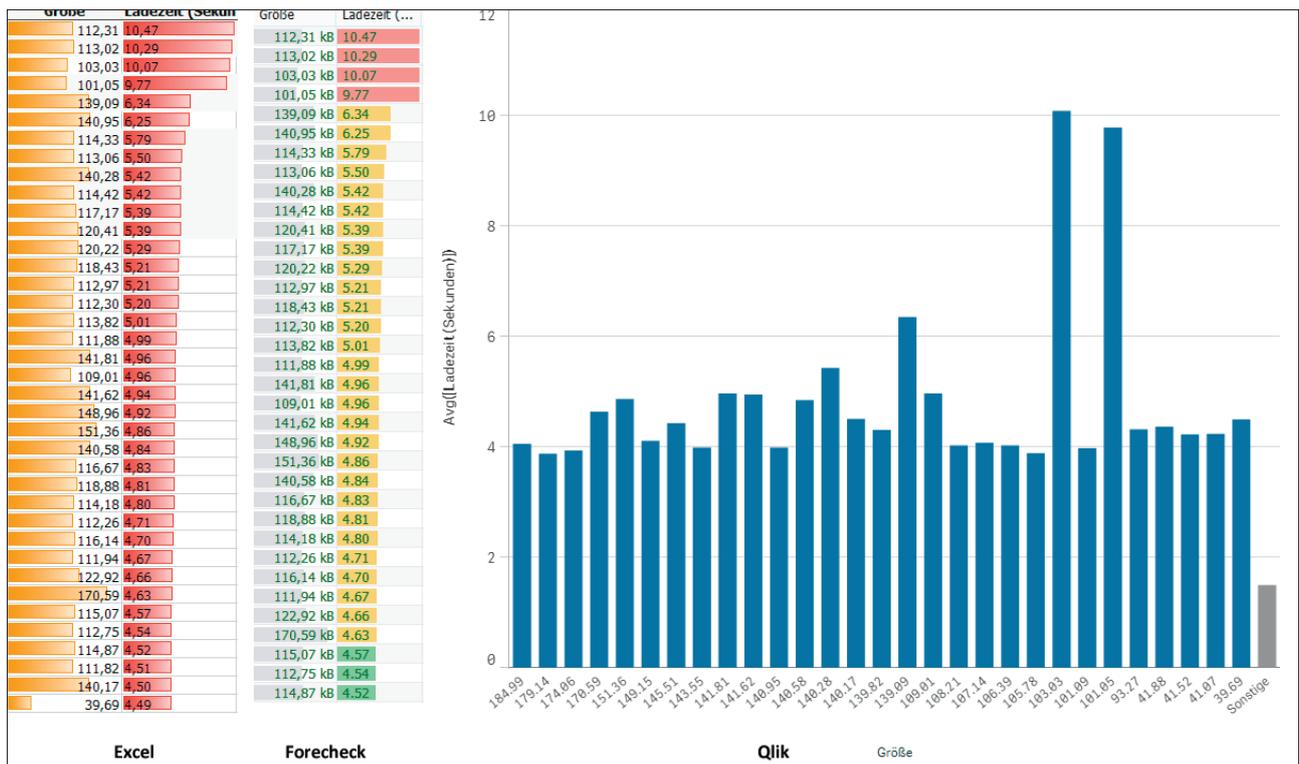


Abb. 13: Vergleich der Datenansicht

Tabelle einen anderen Aufbau hat als die drei Dateien mit den Rankings. Zuletzt fügen Sie noch die Datei Forecheck-sanicare.xls hinzu. Dort sind alle Daten der analysierten Seiten aus Forecheck enthalten. Qlik erkennt, dass sich diese Daten über die Spalte „URL“ mit den anderen verbinden lässt. Das Datenmodell sieht nun aus wie in Abbildung 11. Die ersten importierten Daten sind nicht mit den anderen verknüpft, da es keine Spalten mit gleichem Namen gibt.

Dank der Verknüpfungen können Sie in Diagrammen nun Auswertungen über Datensätze hinweg vornehmen.

Größe vs. Ladezeit

Als Letztes wollen wir viele Daten sinnvoll aggregieren. Erstellen Sie eine neue App und fügen Sie ein Kombidiagramm hinzu. Anschließend importieren Sie die Datei Forecheck-sanicare.xls. Als Dimension verwendet Sie das Feld „Größe“, als Kennzahl das Feld „Ladezeit“ mit der Aggregation

AVG (Ladezeit in Sekunden). Wir wollen wissen, ob die Ladezeit von der Größe einer Seite abhängt. Das Problem ist, dass man in diesem Fall knapp 25.000 Datensätze auswertet. Diese lassen sich schwer in einem einzelnen Balkendiagramm darstellen.

Auf der rechten Seite in Qlik können Sie aber viele Details zu einem Diagramm regulieren. Dort besteht auch die Möglichkeit, eine feste Spaltenzahl einzustellen. Geben Sie dort den Wert 30 ein. Nun fasst Qlik dank der Aggregation automatisch alle Werte zu 30 Balken durch Mittelwerte zusammen.

Die Abbildung 13 zeigt links die bedingte Formatierung in Excel, daneben die bereits eingebauten farblichen Balken bei den Werten Größe und Ladezeit in Forecheck. Rechts sieht man die Auswertung von knapp 25.000 Datensätzen, aggregiert auf 30 Balken in Qlik. Forecheck liefert dank der eingebauten Sortierung der Daten und sinnvoller Farbgebung bereits mehr Hinweise als Excel, doch

nur in der Ansicht von Qlik erkennt man sofort die zwei Ausreißer rund um die Datengröße von ca. 100 kB. Das Beispiel zeigt, dass verschiedene Ansichten tatsächlich auch unterschiedliche „Einblicke“ ermöglichen.

Ausblick

Natürlich zeigt dieser Artikel nur die ersten Möglichkeiten. Für die ersten Schritte mit Qlik empfehle ich Ihnen die beiden Einführungsvideos unter <http://einfach.st/yt8> und <http://einfach.st/yt9>. Die Online-Hilfe zu Qlik finden Sie unter <http://einfach.st/ghelp>. Diese ist sehr umfangreich und ausführlich. Wenn Sie mehr Lust verspüren, können Sie sich auch mal die Software Qlik View ansehen. Für die private Nutzung ist diese ebenfalls kostenlos. Sie bietet noch mehr Möglichkeiten als Qlik Sense, der enthaltene Expressor erinnert stark an die Oberfläche von rapidminder (<http://einfach.st/exp2>). Viel Spaß beim Happy Charting!¶