

Tom Alby

WEBANALYSE: Wie aus Daten Taten folgen

Daten sind das neue Öl, so die Unternehmensberatungen, fast jedes Unternehmen schreibt sich heutzutage „datengetriebene Entscheidungen“ auf die Fahnen. Was auf PowerPoint-Folien und Unternehmens-Flyern wie selbstverständlich klingt, erweist sich in der Praxis im Detail als Herausforderung. In dieser mehrteiligen Serie stellt der bekannte Datenexperte und -analyst Tom Alby ein Framework vor, das den Umgang mit Daten erleichtern soll, erläutert an Webanalyse-Daten, über die fast alle Unternehmen verfügen.

Fast jede Fragestellung, die mit Daten gelöst werden soll, lässt sich in fünf Teilaufgaben untergliedern:

1. Was ist das Problem, das gelöst werden soll?
2. Welche Daten werden benötigt und woher bekomme ich sie?
3. Wie interpretiere ich die Daten richtig?
4. Wie kann ich sicherstellen, dass meine Annahmen richtig sind?
5. Wie kann ich meine Ergebnisse so präsentieren, dass sie verstanden und angenommen werden?

Dies gilt für Data-Science-Themen wie auch für die Webanalyse, vom kleinen Conversion-Optimierungs-Projekt bis hin zum Customer-Churn-Prediction-Modell. Wenn eine der Teilaufgaben nicht korrekt gelöst wird, ist der Rest der Aufgabe mindestens bedroht.

Manchmal werden die Schritte auch durcheinandergebracht. „Meine Absprungrate liegt bei 70 %, das ist viel zu hoch und muss gesenkt werden.“ Abgesehen davon, dass eine Vielzahl der Webanalyse-Anwender die Absprungrate missversteht (Spoiler Alert: Es ist nicht der Anteil der Nutzer, der auf die Seite kommt und „sofort“ wieder geht, dazu mehr in der nächsten Ausgabe), welches Problem wird denn durch eine niedrigere Absprungrate gelöst? Wenn der Datenpunkt schon fehlinter-

pretiert wird, dann wird eventuell ein Problem formuliert, das gar nicht existiert und dessen Behebung auch nicht erfolgreich getestet werden kann. Es kann verführerisch sein, mit den Daten anzufangen und nach Problemen zu suchen (und in manchen Fällen kann das auch gerechtfertigt sein), doch frei nach Ronald Coase: Wenn man Daten lange genug quält, dann werden sie auch irgendwann gestehen.

Erster Schritt: Was ist das Problem, das gelöst werden soll?

Diese Frage stößt nicht selten auf Unverständnis, denn schließlich sollte klar sein, was das Problem ist. Tatsächlich ist es oft nicht klar. Beispiel: Eine Website soll für Benutzer personalisiert werden. Spotify und Amazon zeigen, wie gut Personalisierung funktionieren kann (wenn auch nicht immer), sodass der Wunsch gehegt wird, eine ähnliche Funktionalität auf der eigenen Website anzubieten. Die Webanalyse-Daten sollen helfen, geeignete Vorschläge zu generieren.

Welches Problem wird damit gelöst? Es kann sein, dass dadurch Kaufempfehlungen ausgesprochen werden, die sowieso stattgefunden hätten oder nur einfach vorgezogen werden (eine sogenannte Pull-forward Canni-

Foto: Iuzishan / gettyimages.de

DER AUTOR



Tom Alby ist mehrfacher Buchautor und von der „brand eins“ als „Datenfreak“ bezeichneter Digital-Experte. Nach Stationen wie Google und Ask.com ist er seit 2018 CDO bei Euler Hermes.



Abb.1: Die Messung der Verweildauer in einer Standardinstallation bezieht nicht die letzte Seite ein

balisation). Ebenso könnten Rabatte auf Bundles angeboten werden, aber vielleicht wäre diese Kombi auch so gekauft worden. In diesem Beispiel wird eine vermeintliche Lösung vor ein Problem gestellt, das eventuell gar nicht existiert. Denn um festzustellen, welche Produkte häufig zusammen gekauft werden, muss diese Kombination erst einmal gekauft worden sein, und diese Zusammenstellungen sind auch ohne Hilfe entstanden. Vielleicht sind diese Kombinationen auch rar und ihr Anteil soll erhöht werden. Das Problem ist aber nicht die fehlende Personalisierung. Das Problem ist vielmehr, dass vermutet wird, dass die Besucher nicht so viele Artikel kaufen, wie es wahrscheinlich möglich wäre. Dieses Problem könnte auch anders gelöst werden, denn anstatt darauf zu warten, dass sich Trampelpfade durch Nutzer ergeben, was Zeit kostet, könnten sinnvolle Kombinationen vorab zusammengestellt und empfohlen werden. Der manuelle Aufwand könnte durch frühere Mehrverkäufe kompensiert werden.

Ein Problem hängt meistens mit einem Ziel zusammen und statt der Frage: „Was ist das Problem, das wir lösen wollen?“, könnte auch die Frage gestellt werden: „Welches Ziel wollen wir erreichen?“ Denn ohne ein Ziel ist jede Zahl sinnfrei. Daten sind nur durch Referenzpunkte spannend. Ist eine Conversion-Rate von 1 % gut? Existiert eine Referenz (z. B. Benchmark-Daten), dann kann dieser Datenpunkt eingeordnet werden. Und diese Referenz kann auch ein Ziel sein

(„Es soll eine Conversion-Rate von 2 % erreicht werden“). Liegt die Conversion-Rate darunter, so ist das zu lösende Problem, wie das Ziel erreicht werden kann.

Schwierig wird es, wenn das Ziel entweder nicht vorhanden ist oder spezifisch (ein SMART-Ziel). Geht es nur darum, dass die Awareness gesteigert werden soll, so stellt sich die Frage, in welcher Metrik dies stattfinden soll und ab welcher Zahl dies ausreichend ist. Die benötigte Metrik wird im nächsten Schritt genauer beleuchtet, aber die Metrik ist egal, wenn es kein spezifisches Ziel gibt. So wird der Erfolg von Kampagnen heute noch häufig mittels Anzahl Klicks und Seitenaufrufen gemessen. Aber am Ende des Tages wird ein Problem nicht mit Seitenaufrufen gelöst. Benutzer sollen mit Inhalten interagieren, denn nur dann ist klar, ob es die richtigen Inhalte waren. Ein Seitenaufruf ist die stumpfste Art der Interaktionsmessung, denn er sagt nichts über die Qualität der Interaktion aus. Weniger Seitenaufrufe können sogar gut sein, wenn die verbleibenden Seitenaufrufe zu qualitativ besseren Interaktionen führen. Echte zu lösende Probleme sind zum Beispiel: „Welche Inhalte bringen meine Nutzer dazu, Artikel zu kaufen, sodass ein Umsatzziel erreicht wird?“, oder: „Welche Inhalte bringen meine Nutzer dazu, dass sie wiederkommen und wieder kaufen?“, oder: „Wie kann die Rücksendequote gesenkt werden?“, oder: „Wie kann ich die Qualität der Leads erhöhen, damit das Sales-Team nur noch passende

„Wenn man nicht versteht, wie ein Tool Daten sammelt, sollte man besser die Finger davon lassen.“

Kunden anspricht?“ Die SMART-Ziele müssen gleichzeitig aussagen, ab wann man erfolgreich ist. Die Awareness ist schon gesteigert, wenn ein Nutzer mehr von einer Kampagne weiß. Dass das den Verantwortlichen nicht ausreicht, ist offensichtlich. Aber ohne eine passende Spezifikation der Zielgrößen ist jede Zahl egal.

Zweiter Schritt: Welche Daten werden benötigt?

Google Analytics und Co. bieten bereits in der Standardinstallation viele Daten. Das ist zum einen gut, weil die Einstiegshürde für die Nutzung der Systeme drastisch gesenkt wird. Denn wenn die schlaunen Leute bei Adobe und Google bereits Daten ausgesucht haben, dann werden das sicherlich die Daten sein, die man unbedingt benötigt, oder? Code installieren und schon plätschern die Daten rein, so einfach könnte es sein. Ist es aber leider nicht.

Der einfache Einstieg verdeckt, dass es mit der Standardinstallation in der Regel nicht getan ist. Auch wenn viele Online-Shops existieren, so hat doch jeder seine Eigenheiten, sodass

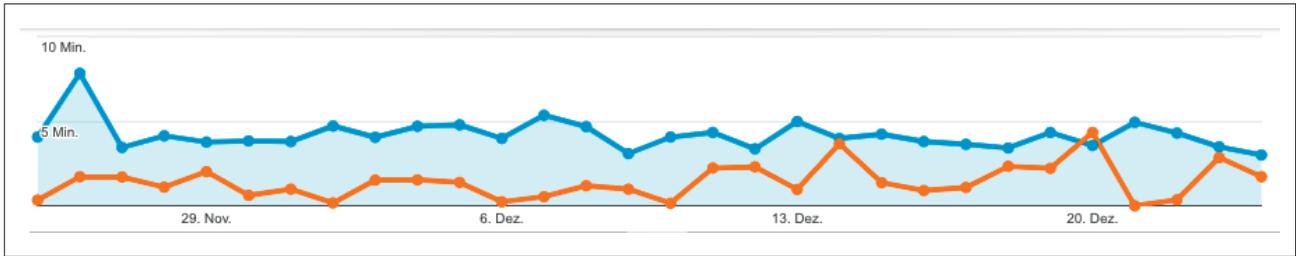


Abb.2: Die Verweildauer vor (orange Linie) und nach einer Anpassung der Messung (blaue Linie)

fast keine Analytics-Installation einer anderen genau gleicht. Google Analytics versucht, den kleinsten gemeinsamen Nenner zu bieten. Mit anderen Worten, es ist nicht unbedingt angebracht, Respekt vor den Daten zu haben, die Google Analytics & Co bereits ab Werk mitbringen, zumal die meisten Daten eh anders interpretiert werden sollten, als es die überwiegende Zahl der Anwender tut.

Hinzu kommt, dass nicht jedem Benutzer klar ist, wie die Daten eigentlich zustande kommen, sei es bei Google Analytics oder anderen Tools, auch abseits der Webanalyse. „A fool with a tool is still a fool“, oder anders gesagt, ein Tool ersetzt nicht das eigene Denken. Es wird ein Tool gesucht, das Suchvolumina abseits vom Google Ads Keyword Planer zur Verfügung stellt, aber wo genau soll so ein Tool die Daten eigentlich herhaben? Manche Tools versprechen sogar, dass sie Daten zum Webtraffic anderer Seiten anbieten können. Woher genau sollen sie diese Daten haben? Schaut man in die Erklärungen auf den Webseiten dazu, findet sich in der Regel nur wenig Sinnvolles. Oder es ist sogar beschrieben, aber so komplex, dass man es nicht versteht. Steigende Kurve bei Google Trends? Bedeutet nicht, dass auch die Suchanfragen gestiegen sind, auch wenn das die Mehrheit der Anwender glaubt (siehe <https://alby.link/googletrends>). Oder die Hilfe wird erst gar nicht gelesen. Zu verstehen, wie Daten genau gesammelt werden, ist ein notwendiger Schritt in den kompetenten Umgang mit Daten.

Ein Beispiel aus der Webanalyse: die durchschnittliche Verweildauer. Um diese zu verstehen, muss zunächst einmal geklärt werden, wie die Verweildauer eigentlich gemessen wird. In einer Standardinstallation der meisten Webanalyse-Tools wird die Verweildauer als zeitliche Distanz zwischen den besuchten Seiten gemessen. Der Nutzer kommt im Beispiel auf Abbildung 1 auf die erste Seite, verbringt dort 30 Sekunden, auf der zweiten Seite dann 20 Sekunden und auf der letzten Seite 40 Sekunden. Nur fehlt dann eine weitere Seite, um die Distanz zu messen. Also wird die auf der letzten Seite verbrachte Zeit nicht gemessen.

Macht nicht viel aus? Da durch diese Art der Messung die Zeit nicht gemessen wird, wenn ein Benutzer nur eine einzige Seite ansieht, kann ein hoher Anteil von Einzelseiten-Besuchen dazu führen, dass ein Großteil der

auf der Website verbrachten Zeit nicht gemessen wird. So ist in Abbildung 2 ein Beispiel zu sehen, wo die Verweildauermessung angepasst wurde, und im Schnitt ergab sich eine fünfmal höhere Verweildauer.

Aber abgesehen davon, dass ein Durchschnitt als Statistik eh problematisch ist (dazu mehr im nächsten Teil), stellt sich auch die Frage, wie sinnvoll die Verweildauer überhaupt ist. Je länger der Text oder je länger ein Video auf einer Seite, desto länger sollte die Verweildauer auf der Seite sein. Aber das würde auch bedeuten, dass diese Informationen auch zur Verfügung stehen während der Analyse. Hinzu kommt, dass, selbst wenn die Messung der Verweildauer angepasst wird, sie nie genau gemessen wird. Die Webanalyse-Systeme bieten somit einen Datenpunkt, der zum einen ungenau gemessen wird, zum andern auch nicht unbedingt handlungsrelevant ist.

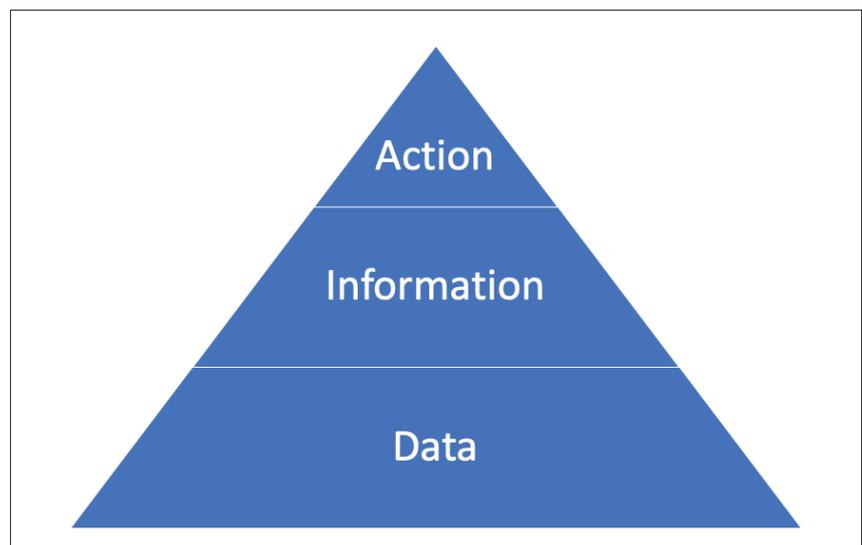


Abb.3: Die Dreifaltigkeit der Daten-Analyse

Handlungsrelevanz ist das Entscheidungskriterium, ob ein Datenpunkt überhaupt benötigt wird. Denn die Dreifaltigkeit der Datenanalyse ist, angelehnt an die Wissenspyramide, die Hierarchie von Daten, Information und Aktion (siehe Abbildung 3). Kann von einem Datenpunkt keine Aktion abgeleitet werden, so wird er nicht benötigt.

Diese Sicht auf Daten ist ungewohnt, denn sie bedeutet, dass nicht alle Daten, die ein Webanalyse-System ab Werk mitbringt, auch benötigt werden. „Die durchschnittliche Verweildauer beträgt 2:06 Minuten.“ Und jetzt? Was kann nun geändert werden? Was wäre ein guter Wert? Viele Berichte kranken genau daran, dass sie nicht verständlich machen, was nun die Aktion aus einem Datenpunkt ist.

Tatsächlich sind die meisten vorgefertigten Dashboards unbrauchbar, wenn nur noch Daten angezeigt werden sollen, aus denen eine Aktion abgeleitet werden kann.

Beispiel: Es soll das Problem gelöst werden, dass die eigene Website zu langsam lädt. Ein guter zu erreichen Wert kann subjektiv festgelegt werden, er kann aber auch anhand von Referenzwerten ermittelt werden. Wenn alle Marktbegleiter-Webseiten in zwei Sekunden laden, die eigene Website aber drei Sekunden benötigt, so ist das ein klarer Wert. Stehen die drei Sekunden alleine da ohne Referenzpunkt, so sind sie wenig handlungsrelevant. Der Datenpunkt drei Sekunden allein ist nutzlos. Mit der Information, dass alle anderen Webseiten eine Sekunde schneller sind, ergibt sich

eine Aktion, nämlich: „Was muss getan werden, damit die eigene Website zumindest genauso schnell ist, wenn nicht sogar schneller?“ Das bedeutet gleichzeitig, dass in einem Dashboard nicht nur die Daten zu der eigenen Seite angezeigt werden, sondern auch die der Marktbegleiter. Diese müssen also zusätzlich erhoben werden. In Google Analytics existieren Daten zur Ladezeit, aber auch hier empfiehlt es sich, genau hinzusehen, wie diese ermittelt werden. Wie im ersten Schritt schon betont: Ohne jeden Bezug ist jede Zahl egal.

Im nächsten Teil geht es dann in der folgenden Ausgabe um typische Fallstricke in der Analyse sowie die Validierung von Hypothesen. ¶

Ultraschnelles
High-Performance
SSD-Webhosting mit nginx