

Dr. Friederike Vial, Markus Schimske, Dr. Pascal Volz

Datengetriebene Attribution bei Air Berlin

Welche Marketing-Kanäle zahlen wie hoch auf die erfolgreiche Akquisition eines Kunden ein bzw. mit welchen Anteilen sind sie für die Conversion verantwortlich? Noch immer wird bei der Beurteilung der Budgetzuteilung auf diese Kanäle nach dem Prinzip „Last Cookie wins“ verfahren. Zwar ist den meisten Marketern klar, dass nicht immer nur das letzte Werbemittel für den Verkauf verantwortlich ist, aber da diese Form einfach zu messen ist, wird sie verwendet – oft mit fatalen Folgen. Ein ökonomisch vernünftigeres, individuelles und datengetriebenes Attributionsmodell aufzustellen, ist sicher etwas aufwendiger. Aber es spiegelt natürlich die Realitäten draußen vor den Bildschirmen deutlich besser wider und verhilft daher zu besseren Entscheidungen und letztlich zwangsweise auch zu mehr Umsatz.

Die Experten Friederike Vial und Markus Schimske von Trakken und Pascal Volz von Air Berlin zeigen auf, wie man die Attribution über die Customer Journey bis zur letztlichen Flugbuchung bei der bekannten Fluglinie mit Google Analytics sehr viel ehrlicher und transparenter abbilden kann.

DIE AUTORIN



Dr. Friederike Vial ist Data Scientist bei Trakken Web Services. Sie ist dort verantwortlich für alle Themen der Datenanalyse und -visualisierung, worunter auch die Berechnung der Regressionsmodelle fällt.

DER AUTOR



Markus Schimske arbeitet bei Trakken Web Services als Consultant für Digitale Analyse. Mit einem Schwerpunkt auf Customer-Journey-Analysen unterstützt er nationale und internationale Kunden bei der Durchführung von Projekten aus diversen Bereichen.

DER AUTOR



Dr. Pascal Volz ist Vice President E-Commerce & Online Sales bei Air Berlin. Vorher war er bei HRS u. a. als Director Performance Marketing mit den verantwortlichen Bereichen SEO, SEA, Display, RTB und Mobile sowie bei der Uni-ster GmbH tätig.

Die Suche nach einem geeigneten Attributionsmodell gestaltet sich in der Praxis nach wie vor schwierig. Zweifelsfrei beginnt ein solches Projekt aber stets mit der genauen Analyse der Customer Journey. Unter der Annahme, dass hierfür alle Voraussetzungen erfüllt sind (vollständiges Tracking aller digitalen Touchpoints mit dem Unternehmen sowie der Conversions), lassen sich erste Erkenntnisse darüber ableiten, wie die „typische“ Customer Journey aussieht. Eine wesentliche Eigenschaft ergibt sich aus der durchschnittlichen Länge der Customer Journey, wobei diese wiederum verschiedene Attribute besitzt. Zum einen definiert sie sich durch die Anzahl der Touchpoints (Sessions) und zum anderen durch die Anzahl der Tage (jeweils vom ersten Kontakt bis zur Transaktion). Hier sind bereits große Unterschiede in Abhängigkeit von der Branche, aber auch von den Produkten zu erkennen. Der Kauf von neuem Druckerpapier wird im Allgemeinen nicht so viel Zeit (gleichbedeutend mit weniger Touchpoints) beanspruchen wie z. B. der Kauf eines Langstreckenflugs für die ganze Familie.

Beide Untersuchungen lassen sich mithilfe der Standardreports des MCF-Bereichs „Time Lag“ sowie „Path Length“ untersuchen. Zu beachten ist dabei allerdings, die Reports richtig zu konfi-

gurieren. Hierbei sollte nicht nur der Conversion-Typ (Goals vs. Transactions), sondern auch das „richtige“ Lookback Window gewählt werden. Das Lookback Window definiert, wie weit Google Analytics in die Vergangenheit des Users zurückschaut, um eine Session zur Customer Journey hinzuzuzählen. Das maximale Lookback Window beträgt in Google Analytics 90 Tage. Das maximale Zeitfenster von 90 Tagen gilt für alle Reports innerhalb der Attributions- und MCF-Berichte von Google Analytics.

Im Fall von Air Berlin zeigte sich über darauf aufbauende Analysen schnell, dass der Zeitraum zwischen dem ersten Kontakt und der Buchung sich häufig proportional zum Flugpreis entwickelt und in vielen Fällen länger als 90 Tage beträgt. Marketing-Kanäle, die klassischerweise am Anfang der Customer Journey stehen, können so nicht mehr „fair“ bewertet werden, da diese in den Reports (z. B. „Top Conversion Paths“) nicht mehr angezeigt werden können. Hiermit kommen wir zum zweiten wesentlichen Bestandteil der Customer-Journey-Analysen: den erfassten Marketing-Kanälen. In der Attribution wird das Ziel verfolgt, Conversions den Marketing-Kanälen so zuzuordnen, dass der tatsächliche Beitrag eines Kanals berücksichtigt wird.

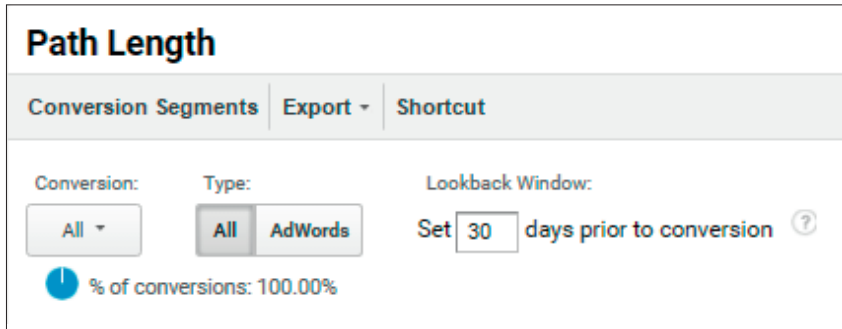


Abb. 1: Konfiguration des MCF-Reports „Path Length“

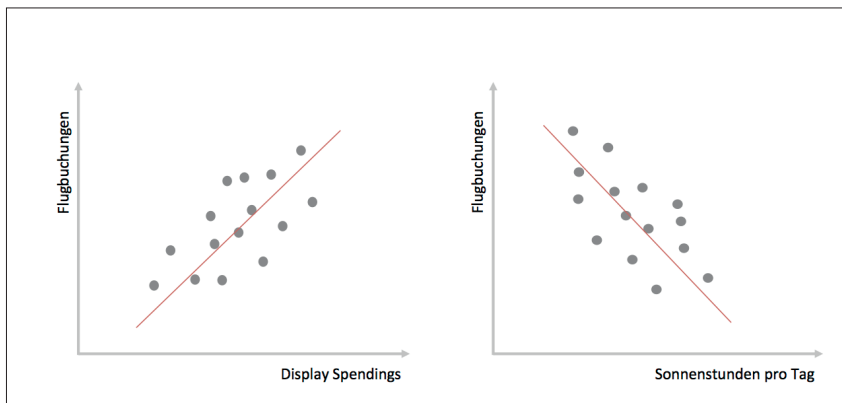


Abb. 2: Positive und negative Zusammenhänge identifizieren und beschreiben

Attributionsmodelle in Google Analytics und ihre Grenzen

Innerhalb von Google Analytics sind bereits die typischen „Standardmodelle“ verfügbar (U-Modell, lineare Verteilung etc.), auf welche hier nicht mehr näher eingegangen werden soll. Sie alle haben eines gemeinsam: Die verwendete Attribution hängt (sofern man dies über ein Custom-Modell nicht individuell anpasst) ausschließlich von der Position des Kanals innerhalb der Journey ab und nicht von dem tatsächlichen (signifikanten) Einfluss. Des Weiteren sind alle Standardmodelle (auch jene mit individuellen Regeln) statischer Natur. Streuverluste, saisonale und klimatische Effekte oder auch ein sich insgesamt veränderndes Umfeld oder die Marktsituation können nicht automatisch berücksichtigt werden. Insbesondere tauchen die Effekte in der Reisebranche auf. Eine ständige (manuelle) Anpassung des Modells wäre somit unvermeidlich.

Für Inhaber einer Google-Analytics-360-Lizenz kann das datengetriebene Modell (DDAM), welches sich alle sieben

Tage automatisch aktualisiert, Abhilfe schaffen. Darüber hinaus hängt die Bewertung eines Kanals dabei nicht von seiner Position, sondern vielmehr von seinem tatsächlichen Beitrag zu einer Conversion ab. Hierzu wertet Google Analytics 360 alle Customer Journeys (egal ob mit oder ohne Conversion) aus, um die Veränderung in der „Conversion-Wahrscheinlichkeit“ für jeden Kanal an jeder Position zu berechnen. Problematisch für Unternehmen mit eher langen Journeys ist aber die Tatsache, dass im DDAM (neben den erneut maximal 90 Tagen Lookback Window) ein weiteres Limit das Ergebnis beeinflusst. Nur die letzten vier Touchpoints vor einer Conversion bekommen einen Anteil an dieser zugesprochen – weitere Touchpoints werden als „Any“ zusammengefasst und nicht weiter berücksichtigt.

Alternative Berechnung mithilfe eines Modellings

Die ökonomische Modellierung bietet einen alternativen Lösungsweg zu den Attributionsmodellen innerhalb von Google Analytics. Die Basis einer

solchen Modellierung bildet die lineare Regression, in welcher die Datengrundlage frei definiert werden kann. Das Verfahren der Regressionsanalyse ist sehr vielseitig und flexibel einsetzbar und ermöglicht nicht nur die Erklärung von Zusammenhängen, sondern auch die Durchführung von Prognosen. Zum Beispiel: Welche Faktoren beeinflussen die Absatzmenge von Schokolade? Sind es die Zutaten wie Kakaogehalt, Zuckeranteil, Gewicht, Nüsse, Nougat und Joghurtcreme oder der Preis pro 100 Gramm, die Jahreszeit oder das Wetter?

Im vorliegenden Fall untersuchten Trakken und Air Berlin den Zusammenhang zwischen den Online-Flugbuchungen und den theoretisch relevanten Einflussfaktoren. Mit der Berechnung ist es sodann möglich, signifikante Zusammenhänge zu identifizieren und den anteiligen Einfluss der einzelnen Marketing-Kanäle und auch der weiteren Faktoren auf die Online-Flugbuchungen zu berechnen. So ließe sich beispielsweise mithilfe einer Modellierung der positive Zusammenhang zwischen den Display Spendings und den Flugbuchungen oder aber der negative Einfluss von Sonnenstunden pro Tag auf die Flugbuchungen belegen und quantifizieren (s. Abb. 2).

In der Realität beeinflussen allerdings deutlich mehr Kanäle die Anzahl der Flugbuchungen. Nicht zuletzt sind neben Spendings und klimatischen Einflüssen auch saisonale Beeinflussungen zu berücksichtigen. Im Rahmen theoretischer Vorüberlegungen wurde ein individuelles Modell für Air Berlin erstellt. Dabei dominierte zunächst die Frage: „Welche Faktoren können die Anzahl der Online-Buchungen beeinflussen?“ Zu beachten ist, dass die Realität mithilfe eines Modells nie komplett abgebildet werden kann. Das heißt, es wird immer Einflussgrößen geben, die im Modell unberücksichtigt bleiben – sei es, weil sie unbekannt sind, weil sie zwar bekannt sind, aber nicht gemessen

werden können, oder weil die entsprechenden Daten nicht zugänglich sind.

Modelldefinition für Air Berlin

Für Air Berlin stellen sich als mögliche Einflussfaktoren zum einen die digitalen Marketing-Aktivitäten und zum anderen die Saison und das Klima heraus. Klassische Marketing-Aktivitäten (TV, Plakat etc.) wurden im Modell bewusst außer Acht gelassen. Die drei definierten Kategorien der Einflussfaktoren können weiter spezifiziert werden (s. Abb. 3). Die digitalen Medien lassen sich zunächst in Kanäle und darüber hinaus weiter aufgliedern: Zum Beispiel ist der Kanal „Paid Search“ weiter zu differenzieren in AdWords und Bing.

Sammlung der Daten

Ein extrem wichtiger und sehr häufig unterschätzter Schritt bei sämtlichen Analysen ist die Datensammlung. Bei der Modellierung gilt dies ganz speziell. Es ist auf eine gute Datenbasis zu achten, denn auf dieser Grundlage wird das Modell berechnet. Ist die Datenbasis nicht einwandfrei, so kann daraus weder ein korrektes Modell noch eine exakte Prognose resultieren.

Eine erste Entscheidung muss bei der Datensammlung bezüglich des Analysezeitraums getroffen werden. Hier gilt es, einen eher längeren Analysezeitraum zu wählen, denn dieser ermöglicht trotz vieler Einflussvariablen den Nachweis kleiner und mittlerer Effekte. Außerdem können nur so die saisonalen Auswirkungen nachgewiesen werden. Bei Air Berlin wurden die historischen Daten mehrerer Monate als Input für die statistischen Berechnungen genutzt, aus denen schließlich das Modell hervorging. Die Daten wurden auf Tagesbasis gesammelt und ausgewertet. Somit sind die Ergebnisse für die Budgetplanung auf Tagesbasis nutzbar.

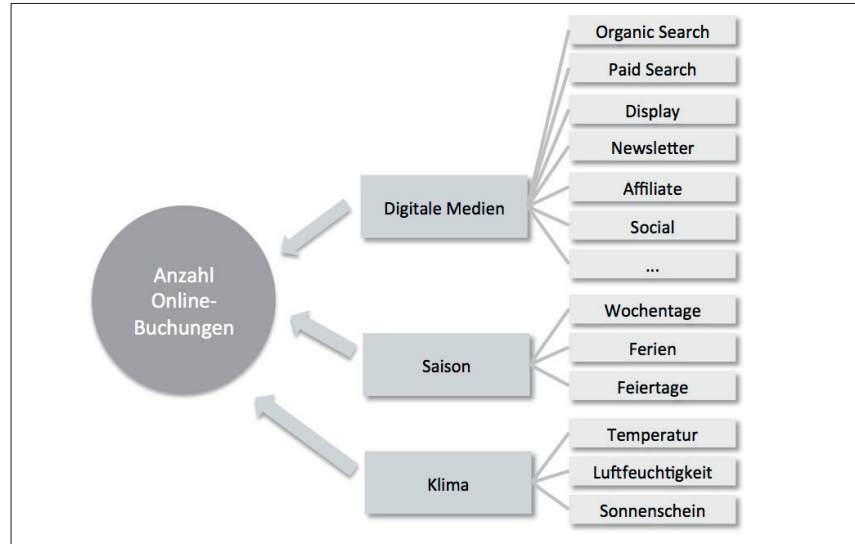


Abb. 3: Einflussfaktoren für die Flugbuchungen auf airberlin.com

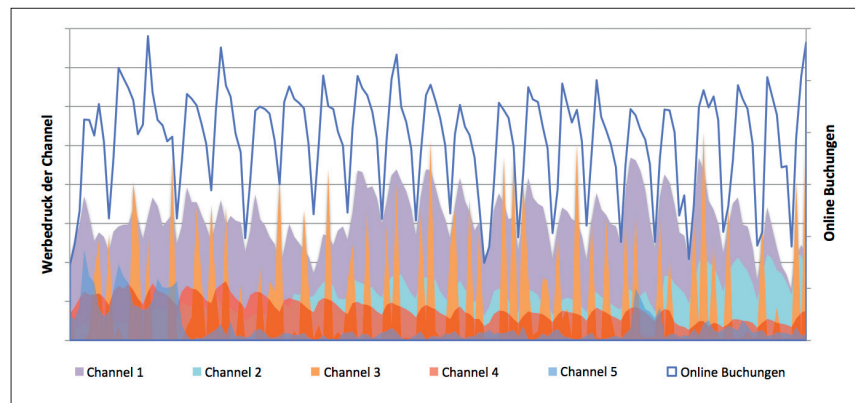


Abb. 4: Zeitliche Entwicklung der Flugbuchungen und Impressions pro Marketing-Kanal (im Analysezeitraum)

Neben der Bestimmung des Analysezeitraums sind die Datenquellen zu benennen. Theoretisch kommen als Input verschiedene Metriken pro Marketing-Kanal infrage: zum Ersten die Anzahl der Impressions pro Kanal, zum Zweiten die Spendings pro Kanal und zum Dritten die Anzahl der Sessions, die über einen bestimmten Kanal generiert wurden. Im Fall von Air Berlin wurden die Impressions auf Tagesbasis als Metrik in der Modellierung verwendet. Hieraus folgte, dass diverse Quellen bei der Datensammlung zum Einsatz kamen. Google Analytics als alleinige Datenquelle genügt bei dieser Vorgehensweise nicht. Dieses Tool wurde lediglich genutzt, um die Tagesdaten der Zielvariable „Anzahl der Online-Flugbuchungen“ zu exportieren.

Kennenlernen der Daten, indem sie beschrieben werden

Im Rahmen der deskriptiven Analyse werden die Daten vor der Modellierung zunächst beschrieben: Welche Kanäle veränderten sich im Betrachtungszeitraum wie? Welche Kanäle besitzen einen großen Anteil am gesamten Werbedruck? Die Analyse zeigt, welche Kanäle in welchem Umfang in der Vergangenheit gebucht wurden. Eine Auswertung für den gesamten Analysezeitraum ermöglicht außerdem die Darstellung der Entwicklung einzelner Kanäle im direkten Vergleich mit der Entwicklung der Online-Buchungen. Wie Abbildung 4 zeigt, sind zwar Muster zu erkennen, eindeutige Abhängigkeiten können allerdings nicht identifiziert werden. Hierfür ist es notwendig,

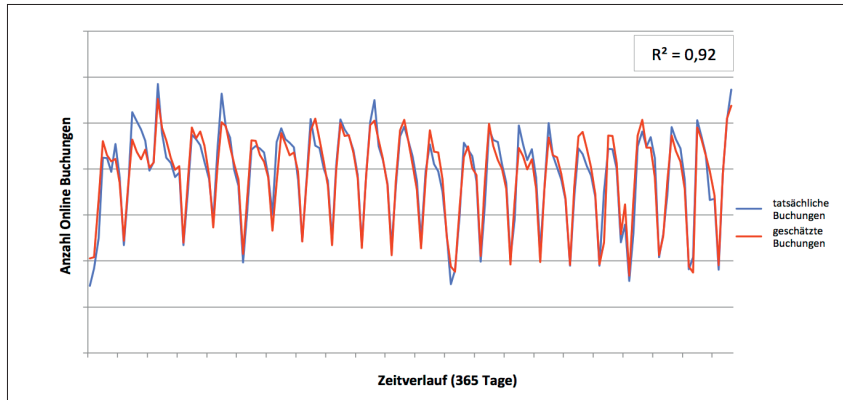


Abb. 5: Modellgüte, d. h. Vergleich der tatsächlichen und der geschätzten Flugbuchungen

die Regressionsanalyse durchzuführen.

Für die Durchführung der Regressionsanalyse kommen verschiedene Statistikprogramme infrage, wie z. B. SPSS, R, Stata oder auch EViews. Die Berechnungen des Air-Berlin-Modells wurden in R durchgeführt, nachdem die Daten für die Analyse zusammengeführt und aufbereitet worden waren. In mehreren Berechnungsschritten wurde sodann anhand diverser Kenn-

zahlen das optimale Modell für Air Berlin schrittweise zusammengestellt. Um es schließlich zu bewerten, wird zunächst betrachtet, inwieweit die geschätzten Flugbuchungen den tatsächlichen Flugbuchungen entsprachen. Dieser Vergleich ist in Abbildung 5 dargestellt: Das für Air Berlin berechnete Modell kann aufgrund des hohen Bestimmtheitsmaßes von 0,92 mit sehr gut bewertet werden. Und auch die grafische Darstellung zeigt, dass die

beiden Werte „tatsächliche Buchungen“ und „geschätzte Buchungen“ i. d. R. nahe beieinanderliegen. Ein weiterer Check, der sogenannte Hold-out-Test, bestätigt darüber hinaus, dass das Modell auch für Prognosen geeignet ist.

Mit der ökonomischen Modellierung konnte die signifikante Wirkung der untersuchten Kanäle und weiterer Einflussgrößen sowie deren jeweilige Wirkungsstärke berechnet werden. Bei der zeitlichen Betrachtung der sogenannten Dekomposition (s. Abb. 6) sieht man nicht nur, welche Wirkung von den signifikant einflussreichen Kanälen auf die Anzahl der Flugbuchungen ausgeht, sondern auch, wie sich diese Wirkung im betrachteten Zeitraum veränderte. Die veränderte Wirkung ist auf eine neue Online-Marketing-Strategie innerhalb des Analysezeitraums zurückzuführen. So nimmt beispielsweise die anteilige Wir-

WE LOVE BOOSTING



ABO
+ EMAILLE-TASSE
GESCHENKT!

AKTION! inkl. Emaille-Tasse



JETZT ABONNIEREN!

Einfach und schnell unter www.websiteboosting.com/abo

- » 6 Ausgaben zum Vorteilspreis von nur EUR 51,00 (D)**
- » inkl. exklusiver limitierter Emaille-Tasse
- » Jeder Titel wird direkt zu Ihnen nach Hause geliefert, ohne zusätzliche Zustellkosten.
- » immer ein paar Tage früher informiert sein

** Sie sparen 7,80 EUR im Vergleich zum Einzelkauf. Im Ausland: EUR 63,00 inkl. Versandkosten

Mit dem Smartphone scannen oder



www.websiteboosting.com/abo

kung von Channel 2 im Laufe der Zeit zu.

Eine Berechnung der Wirkungsanteile unabhängig vom Zeitpunkt verdeutlicht außerdem, welche Kanäle insgesamt einen hohen Erklärungsanteil an den Flugbuchungen haben. Es kann damit eine Rangordnung der einflussreichen Marketing-Kanäle erstellt werden. Gleichzeitig ist es möglich, die Wirkungsanteile mit den Anteilen bezüglich der Impressions zu vergleichen. Im Fall von Air Berlin fiel beispielsweise die starke Wirkung eines Kanals im Modell auf, der an den gesamten Impressions lediglich einen sehr geringen Anteil aufwies (s. Abb. 7).

Außerdem ist es mit dem Modelling möglich, die unterschiedlichen Nachwirkzeiten für die Marketing-Kanäle zu bestimmen. Hiermit kann festgestellt werden, ob es sich jeweils um einen initialen, assistierenden oder abschließenden Kanal handelt (s. Abb. 8). Es ist somit möglich, die Customer Journey anhand der signifikant einflussreichen Kanäle zu beschreiben. Einen weiteren wertvollen Hinweis gibt die Modellierung mit Blick auf die Art des Zusammenhangs. Mit der Information, ob ein linearer Zusammenhang oder ein Zusammenhang mit abnehmendem Grenznutzen vorliegt, können künftige Budgets noch besser datengetrieben verteilt werden. Dadurch können ganz konkret Kanäle bestimmt werden, für die sich eine Budgeterhöhung anbietet.

Operationalisierung der Ergebnisse

Die Ergebnisse der Modellierung haben viele Insights generiert. Besonders relevant sind dabei die Erkenntnisse zu der anteiligen signifikanten Wirkung einzelner Kanäle auf die Flugbuchungen, zur Nachwirkzeit und zur Art des Wirkungszusammenhangs. Air Berlin führte im Anschluss an das Modelling eine Anpassung der Allokation des Online-Budgets durch. So wurde der Fokus für Veränderungen zunächst auf Channel 5 gelegt, da die Analyse einen hohen Erklärungsanteil

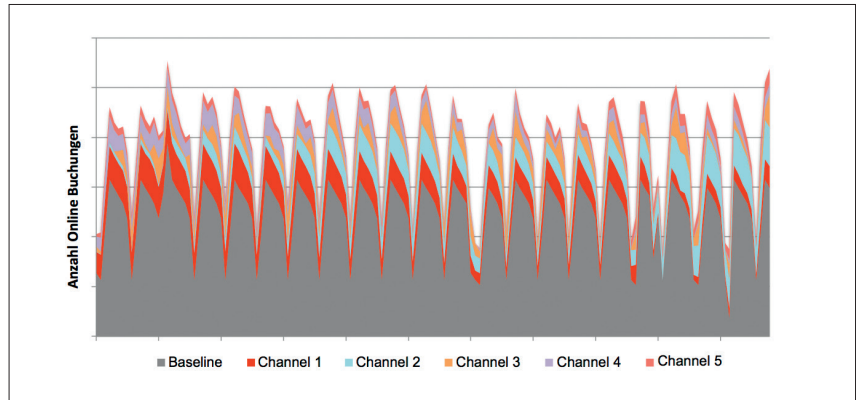


Abb. 6: Einflussreiche Marketing-Kanäle und deren anteilige Wirkung im Zeitverlauf

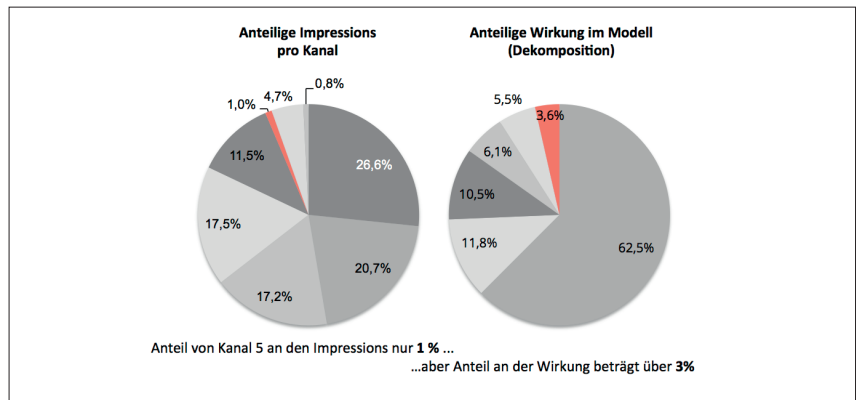


Abb. 7: Vergleich der Anteile je Marketing-Kanal (Impressions vs. Wirkung im Modell)

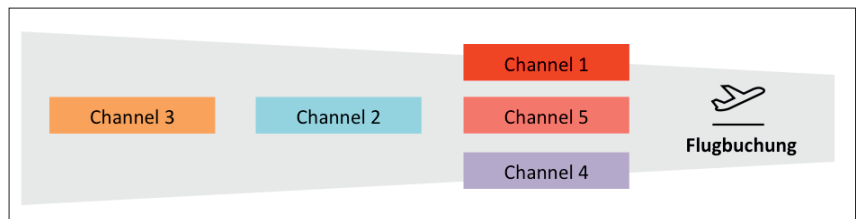


Abb. 8: Position der Marketing-Kanäle innerhalb der Customer Journey

an den Flugbuchungen belegt, obwohl der Anteil an den gesamten Impressions vergleichsweise gering war. Bei den Überlegungen zur optimalen Budgetallokation flossen auch die Kosten pro Kanal ein, denn nur mit einer monetären Beurteilung wird ein umfassendes Bild gezeichnet. Darüber hinaus wurde für Channel 5 ein linearer Zusammenhang zwischen den Impressions und den Flugbuchungen identifiziert, woraus gefolgert werden kann, dass sich eine Budgeterhöhung direkt auf die Flugbuchungen auswirken wird. Mit diesen Informationen konnte anschließend ein Experiment gestartet werden, bei dem das Budget für Channel 5 angehoben wurde. Mithilfe einer weiteren Regressionsmodellierung konnte der

Erfolg dieser Budgetveränderung nachgewiesen werden.

Fazit

Zusammenfassend kann festgehalten werden, dass die Berechnung eines Regressionsmodells eine gute Alternative der Attribution zur Lösung innerhalb von Analytics darstellt. Allerdings gibt es bei diesem Vorgehen natürlich auch Grenzen. So muss zum Ersten ein höherer zeitlicher Aufwand für die Datensammlung und -harmonisierung eingeplant werden. Zum Zweiten finden die Berechnungen teilweise manuell statt, wodurch die Ergebnisse nicht automatisch und vor allem nicht so schnell vorliegen wie in Google Analytics. ¶