

Autor: Bora Beran, PhD

R und visuelle Analyse – Leistungsfähigkeit

Zusammenspiel von komplexen
Statistiken und Datenvisualisierung

Datenforscher und Statistiker kennen das leistungsstarke Analysetool unter der schlichten Bezeichnung „R“. R ist ein Open-Source-Produkt, eine freie Programmiersprache und Umgebung für statistische Berechnungen. Sie zählt zu den am schnellsten wachsenden statistischen Programmiersprachen und gilt für immer mehr Datenforscher als das Tool ihrer Wahl. Genau genommen liegt R inzwischen vor SPSS Statistics und SAS und ist das von Datenforschern am häufigsten verwendete Tool, das inzwischen von etwa der Hälfte aller Datenforscher eingesetzt wird.^{1 2}

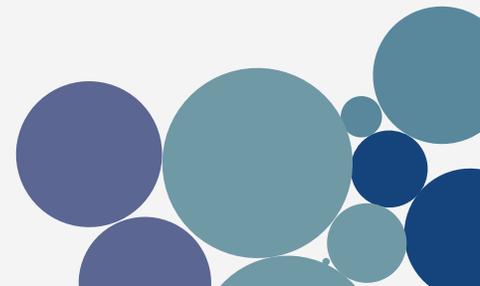
Statistiker, Datenforscher und Analysten werden immer mehr ein integraler Bestandteil einer Organisation, deren Aufgabe darin besteht, die Unmengen der heute verfügbaren Daten zu sichten und wertvolle neue Erkenntnisse zu gewinnen. Ganz gleich, ob es darum geht, komplexe Muster und Beziehungen zu identifizieren oder die Daten für die Prognose künftiger Maßnahmen einzusetzen: Um diese Erkenntnisse aufzudecken, bedarf es der Fähigkeit, statistische Modelle zu erstellen und zu analysieren sowie die Ergebnisse mit anderen zu teilen – genau die richtige Aufgabe für die R-Integration in Tableau.

Während R die „Knochenarbeit“ der statistischen Auswertung übernimmt, ermöglicht Ihnen Tableau, die Ergebnisse in einer einfachen Drag-and-Drop-Umgebung zu untersuchen und zu visualisieren. Dabei werden Ihre statistischen Ergebnisse in ein ansehnliches Dashboard umgewandelt. Mit Tableau Server können Sie diese Visualisierungen im gesamten Unternehmen gemeinsam und sicher nutzen, indem Sie Fachanwendern die schwierigen Aufgaben der Datenforscher anvertrauen. Wenn sich diese Ergebnisse bequem einsehen und mitteilen lassen, können diese Auswertungen auch eine wesentlichere Rolle bei der Entscheidungsfindung innerhalb der gesamten Organisation einnehmen.

Zusammen bieten Tableau und R das Beste aus beiden Welten, weil sich Daten nun schnell und einfach nutzen lassen, um die heute immer komplexer werdenden Herausforderungen einer Organisation zu meistern.

Inhaltsverzeichnis

Licht ins Dunkel bringen – was Kunden meinen, wenn sie etwas sagen	3
Entscheidungsfindungsprozesse automatisieren.....	4
Große, komplexe Datensätze verstehen	5
Beweisen, dass es kein Zufall war.....	5
Der Autor	6

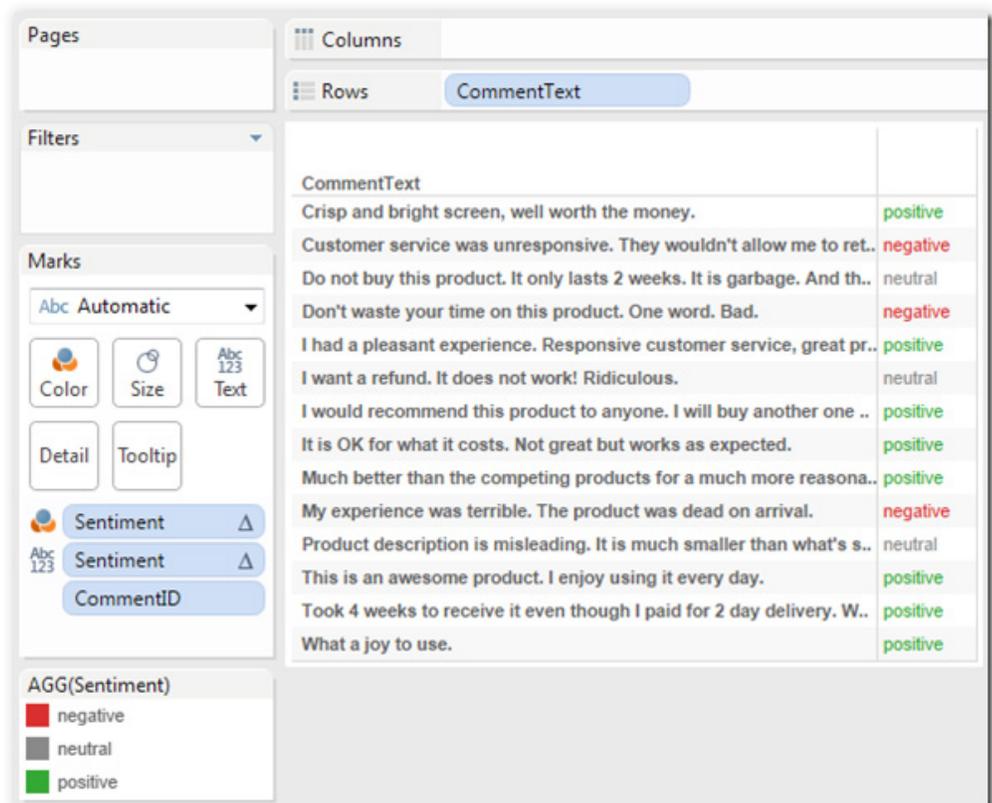


Licht ins Dunkel bringen – was Kunden meinen, wenn sie etwas sagen

Sentiment-Analyse

Online-Meinungen wie Kommentare, Bewertungen und Diskussionen auf beliebten Websites wie Twitter, Facebook, YouTube, Pinterest und Amazon stellen für Unternehmen, die ihre Kundenbeziehungen verwalten möchten, eine Fundgrube verborgener Schätze dar. Um die Juwelen jedoch hervorholen zu können, muss Licht ins Dunkel gebracht werden. Von der Produktentwicklung bis hin zum Reputationsmanagement müssen Organisationen die Online-Konversationen systematisch durchforsten, um sie zu verstehen und Maßnahmen ergreifen zu können. Hierzu entscheiden sich viele für die Sentiment-Analyse.

R bietet eine Funktionalität, mit der Texte und Stimmungen leichter analysiert werden können. Gleichwohl müssen Benutzer sich den Output ansehen können, um ihn für die Entscheidungsfindung zu nutzen. Durch die R-Integration in Tableau können Sie diesen Output in eine Visualisierung übertragen, in der positive und negative Kommentare farblich sichtbar gemacht werden.



Bildunterschrift: Durch die Visualisierung der Ergebnisse aus der Sentiment-Analyse mit R lassen sich Kundenprobleme leicht erkennen.

Im Handumdrehen bringt die Sentiment-Analyse Licht ins Dunkel großer Datenmengen und zeigt dem Geschäftsanwender auf, was Kunden meinen, wenn sie etwas sagen. Hiermit kann Ihr Unternehmen spezifische Themen oder sich verändernde Trends rasch identifizieren und mit einer gezielten Öffentlichkeitsarbeit und Marketingstrategie darauf reagieren.

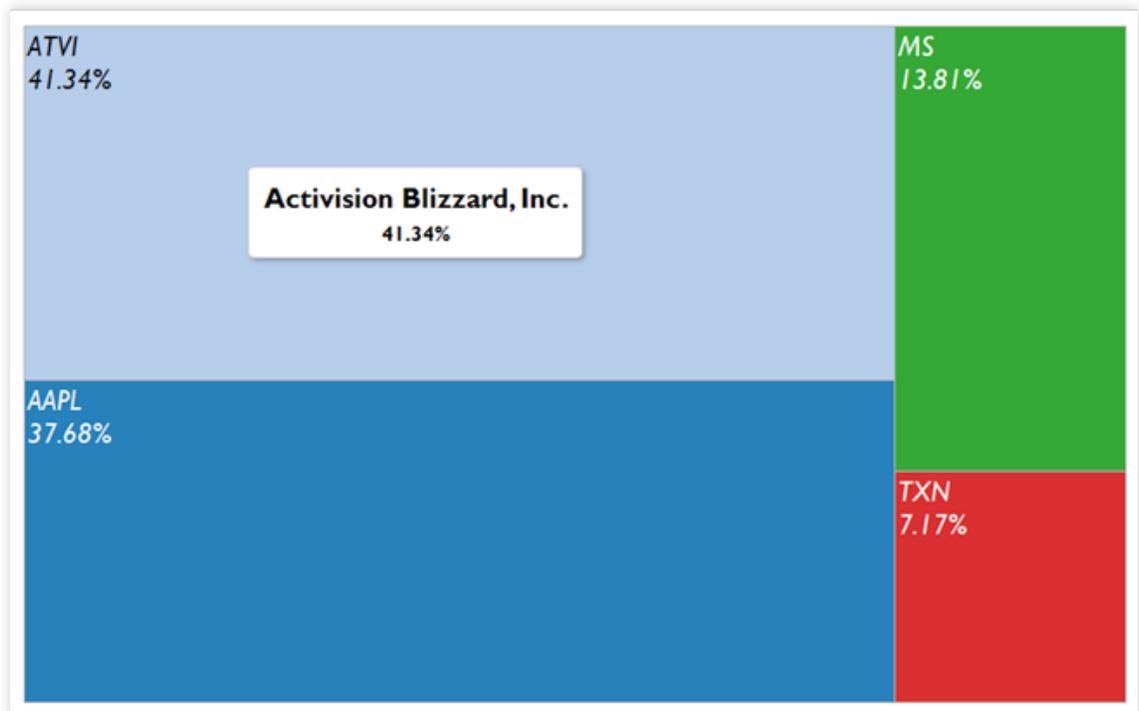
Entscheidungsfindungsprozesse automatisieren

Präskriptive Analytik

Haben Sie sich schon einmal gefragt, wie Fluggesellschaften eigentlich ihre sich ständig ändernden Ticketpreise festlegen? Fluggesellschaften sichten eine komplexe Reihe von Reisevariablen, Kundenanfragen und Zeitabläufen, um in Echtzeit Ticketpreise anzubieten, die zum Kauf verlocken, ohne jedoch die Gewinnmarge zu drücken. Unter Vorgabe bestimmter Rahmenbedingungen suchen sie also nach dem optimalen Ticketpreis. Hierin liegt die Aufgabe der präskriptiven Analytik.

Während die deskriptive Analyse beschreibt, was bereits geschehen ist und die prädiktive Analyse das beschreibt, was geschehen könnte, hilft die präskriptive Analytik bei der Entscheidung, was geschehen sollte. Angesichts mehrerer Entscheidungsmöglichkeiten wird die Auswertung in Hinblick auf das bestmögliche Ergebnis vorgenommen. Die präskriptive Analytik lebt von großen Datenmengen.

R handhabt die präskriptive Analytik recht unkompliziert durch den Einsatz von Optimierungsfunktionen. Mit Tableau können Sie diese Ergebnisse dann visualisieren und kommunizieren, so dass die Organisation Entscheidungen auf Grundlage der statistischen Arbeitsleistung von R treffen kann. So finden Sie beispielsweise unten ein Beispiel für den in einer Baumkarte veranschaulichten `optim()` Output in R. In diesem Fall wird die optimale Portfolio-Allokation für vier Aktien visualisiert.



Bildunterschrift: Tableau erleichtert die Visualisierung der Ergebnisse aus der präskriptiven Analyse in R, wie dieses Beispiel einer Portfolio-Allokation zeigt.

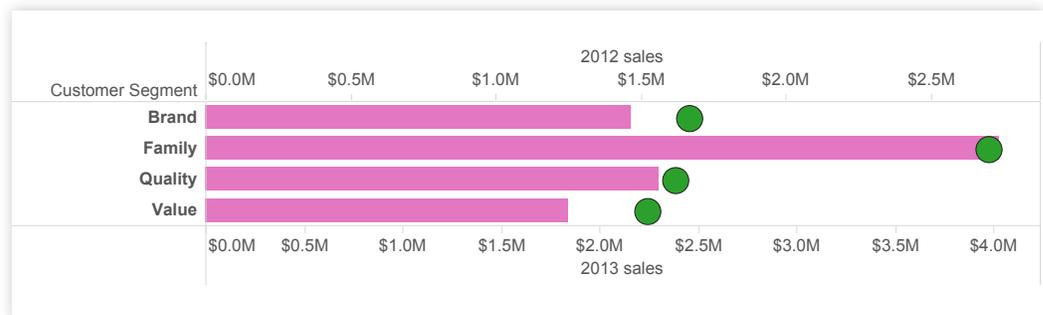
Große, komplexe Datensätze verstehen

Entscheidungsstrukturen

Manchmal sind Daten derart komplex, dass es schwierig ist, ein umfassendes Regressionsmodell oder ein einzelnes prädiktives Schema zu finden, mit dem sich alles beschreiben lässt. Eine Lösung besteht darin, Ihre Daten so lange in kleinere Untergruppen zu unterteilen, bis sich jede Untergruppe anhand eines einfachen Modells erklären lässt. R ermöglicht es dem Benutzer, rasch einen Entscheidungsbaum zu erstellen, um komplexe Daten zu klassifizieren.

Hat Ihr Team einmal ein R-Skript für die Klassifizierung von Daten unter Verwendung eines Entscheidungsbaums geschrieben, lässt sich diese Klassifizierung bei jeder Datenvisualisierung verwenden, um die Farbe, Form oder Gruppierung bequem per Drag-and-Drop zu bestimmen. Genau genommen kann jeder Geschäftsanwender seine Ergebnisse in einer Datenvisualisierung nutzen, wenn das R-Skript zur Datenklassifizierung einmal erstellt worden ist.

In der Visualisierung unten haben wir einen Entscheidungsbaum zur Klassifizierung der Kunden einer Handelsmarke auf Grundlage ihres Kaufverhaltens erstellt. Anschließend haben wir diese Klassifizierung in einem einfachen Balkendiagramm dargestellt, wo die Umsätze für diese Gruppen aus den Jahren 2012 und 2013 einander gegenübergestellt werden. Durch den Einsatz von Tableau mit R kann somit jeder Geschäftsanwender aufbauend auf der Entscheidungsbaum-Funktion, die in R von einem Datenforscher erstellt wurde, Daten untersuchen. Auf diese Weise wird es Geschäftsanwendern möglich, täglich auf die aussagekräftige Arbeit eines Datenforscherteams zuzugreifen.



Bildunterschrift: Diese Visualisierung zeigt Klassifizierungen von Kundensegmenten, die mit einem Entscheidungsbaum in R erstellt worden sind. Der Geschäftsanwender kann diese Daten bequem mit Umsatzdaten kombinieren (in diesem Fall Daten aus 2012 und 2013), um nachzuvollziehen, inwieweit die Ausgaben eines jeden Segments zwischen den einzelnen Jahren schwanken.

Beweisen, dass es kein Zufall war

Statistische Prüfverfahren

Ebenso wie vor Gericht, wo die Unschuldsvermutung eines Beschuldigten solange gilt, bis seine Schuld bewiesen ist, zielt auch das statistische Prüfverfahren darauf ab, den Nachweis zu erbringen, dass etwas wahr ist, indem man ausschließt, dass es unwahr ist. In den meisten Fällen bedeutet das, nachzuweisen, dass etwas nicht bloß rein zufällig geschehen ist. Dies bezeichnet man häufig als „statistisch signifikant“.

So versucht man beispielsweise in der pharmazeutischen Forschung nachzuweisen, dass ein bestimmtes Medikament etwas bewirkt, dass also die Ergebnisse für den Erfolgsnachweis statistisch relevant waren und nicht bloß auf Zufällen beruhen.

Ihre Stichprobe kann grundsätzliche Daten wie Geschlecht oder ethnische Zugehörigkeit oder auch Maßeinheiten wie Gewicht, Alter oder sogar Zeitspannen umfassen. Anhand dieser Daten möchten Sie möglicherweise die Wahrscheinlichkeit untersuchen, dass ein spezielles Arzneimittel bei Mitgliedern einer bestimmten Altersstruktur im Vergleich zu jenen der Kontrollgruppe eine positive Wirkung zeigt.

R bietet eine Vielzahl an Funktionen für statistische Prüfverfahren, während Tableau die Ergebnisansicht dieser Tests erleichtert. Dank der Parameter von Tableau lassen sich in Prüfverfahren die Optionen wie beispielsweise der Zwei-Stichproben-t-Test nach Welch vom Benutzer bequem anpassen und die Ergebnisse visualisieren.



Bildunterschrift: Tableau erleichtert die Visualisierung der Ergebnisse aus den Funktionen der statistischen Prüfverfahren in R und verfügt über Filter, mit denen der Benutzer die Optionen des Prüfverfahrens anpassen kann.

Der Autor

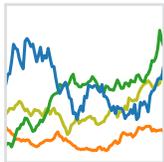
Bora Beran ist Programm Manager bei Tableau Software. Sein Team befasst sich schwerpunktmäßig mit Statistik & Berechnungen, technischen Partnerschaften und Abfragegenerierung. Bevor er zu Tableau kam, war Bora Beran bei Microsoft Research, SQL Server und Windows HPC Server tätig und arbeitete an Tools und Infrastrukturen für umfangreiche Berechnungsmodelle und Simulationen, Data Mining, Visualisierung und Wissensrepräsentation. Er promovierte an der Drexel Universität und ist Lektor bei verschiedenen wissenschaftlichen Fachzeitschriften im Bereich Technical Computing.

¹Karl Rexer, Heather Allen, & Paul Gearan (2011); [2011 Data Miner Survey Summary](#), vorgestellt auf der Predictive Analytics World, Okt. 2011.

²Karl Rexer, Heather Allen, & Paul Gearan (2010); [2010 Data Miner Survey Summary](#), vorgestellt auf der Predictive Analytics World, Okt. 2010.

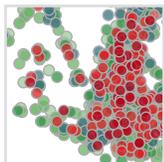
Informationen zu Tableau

Tableau Software hilft Menschen dabei, Daten sichtbar und verständlich zu machen. Mit Tableau können Nutzer Informationen schnell analysieren, visualisieren und mit anderen teilen. Mehr als 19.000 Kunden nutzen Tableau für schnelle Analysen im Büro und unterwegs. Zehntausende Nutzer verwenden Tableau Public, um anderen Daten in Blogs und auf Websites zur Verfügung zu stellen. Um zu erfahren, welche Vorteile Tableau für Sie hat, laden Sie die kostenlose Testversion unter www.tableausoftware.com/trial herunter.



Weitere Materialien

[Kostenlose Testversion herunterladen](#)

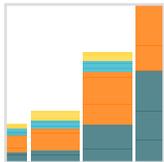


Weitere Literatur zum Thema

[Das R-Projekt für statistische Berechnungen \(R Project for Statistical Computing\) ansehen](#)

[Erfahren Sie mehr über die R-Integration bei Tableau](#)

[Bora Berans Blog ansehen](#)



Weitere Quellen erkunden

- [Produkt-Demo](#)
- [Schulungen & Lernprogramme](#)
- [Community & Support](#)
- [Kundenberichte](#)
- [Lösungen](#)

