

Kundensegmentierung

Individuelle Kundenansprache trotz optimaler Ressourcennutzung

Dr. Steffen Wagner



Ihre Kunden sind Individuen. Sie haben ganz unterschiedliche Bedürfnisse, finanzielle Hintergründe und Kaufverhaltensweisen. Welchen Schluss müssen Sie als Unternehmen aus dieser Tatsache ziehen? Sie können Ihre Kunden nicht alle über einen Kamm scheren und mit ihnen umgehen, als handele es sich um eine homogene Gruppe. Das ist natürlich leichter gesagt als getan. Denn die individuelle Ansprache eines jeden Kunden ist zeit- und kostenintensiv. Man muss nicht BWL studiert haben, um zu wissen, dass Kosten und Nutzen in einem solchen Szenario schwer unter einen Hut zu bringen sind. Die Lösung ist ein gelungener Mittelweg: eine Kundensegmentierung. Ihre Kundenbasis wird dabei mithilfe statistischer Methoden in eine handhabbare Anzahl homogener Untergruppen oder Segmente eingeteilt. Ein Segment umfasst Kunden, die sich aufgrund ihrer Eigenschaften und Bedürfnisse sehr ähnlich sind. Eine segmentspezifische Ansprache erlaubt es, Ressourcen zu bündeln und gleichzeitig eine Individualisierung der Kundenbeziehung zu erreichen. Solche Gruppen können beispielsweise Schnäppchenjäger, Fashionistas oder Technikfreaks sein. Werbe- und CRM-Maßnahmen lassen sich dann segmentspezifisch aussteuern, während Ressourcen betriebswirtschaftlich optimal eingesetzt werden. Oder kurz gesagt: Sie erreichen eine Win-win-Situation.

Das vorliegende Whitepaper informiert umfassend zum Thema Kundensegmentierung. Voraussetzungen, Ablauf und Einsatzmöglichkeiten eines solchen Projekts werden diskutiert und die wesentlichen Punkte beleuchtet.

Zielstellung der datengetriebenen Kundensegmentierung

Eine Kundensegmentierung bietet die Möglichkeit, datengetrieben homogene Kundengruppen zu identifizieren, die sich z.B. in Bezug auf Kaufverhalten und andere Eigenschaften sehr ähnlich sind, während sich die Gruppen untereinander möglichst stark unterscheiden. Abbildung 1 fasst die Ziele einer datengetriebenen Kundensegmentierung zusammen.

Insights über Ihre Kundenbasis

Bestimmt existieren auch in Ihrem Unternehmen zahlreiche Hypothesen von *typischen* Kunden oder Personas, die aus regelmäßigen Erfahrungen und Beobachtungen im Arbeitsalltag resultieren. Eine datengetriebene Kundensegmentierung deckt auf, welche dieser hypothetischen Segmente sich tatsächlich in den Daten wiederfinden lassen und auch, wie groß deren Anteile an der Kundenbasis sind. Wir haben oftmals die Erfahrung gemacht, dass bei der Analyse Segmente zutage treten,

die bisher so noch nicht wahrgenommen oder erwartet wurden.

Individualisierte Ansprache

Wir wissen, dass Kunden unterschiedlich sind und entsprechend unterschiedlich behandelt werden sollten. Kundensegmente bieten die Möglichkeit homogene Gruppen individualisiert anzusprechen, so kann z.B. ein segmentspezifisches Newsletter-Konzept entworfen werden. Spezifischer und relevanter Content vermeidet Irritationen beim Empfänger, da sich diese durch das werbende Unternehmen entsprechend ihrer Charakteristika wahrgenommen fühlen. Auch der Zeitpunkt der Ansprache kann an die Eigenschaften des Segments angepasst werden, wenn sich z.B. ein Segment durch vermehrte Bestellungen zu besonderen Anlässen auszeichnet. Einzelne unrentable Kundengruppen werden möglicherweise sogar ganz von der Ansprache ausgeschlossen.

Datengetrieben statt Hypothesenbasiert

Prinzipiell lässt sich eine Kundensegmentierung auch durch zuvor festgelegte Heuristiken konzipieren. So könnte man bestimmte Regeln in Bezug auf Alter, Geschlecht und Bestellverhalten formulieren, nach denen dann Kunden den Segmenten zugeordnet werden. Das Ergebnis kann jedoch nur so gut sein wie die Übereinstimmung zwischen Regelwerk und Kundencharakteristika. Passen die Regeln nicht zur Kundenbasis, besteht die Gefahr, dass die heuristisch festgelegten Gruppen nicht trennscharf und in sich homogen sind. Erfolgt die Segmentierung hingegen datengetrieben, so ist sichergestellt, dass die identifizierten Segmente zur tatsächlich vorhandenen Kundenstruktur passen.

Optimaler Einsatz von Ressourcen

Nicht zuletzt ermöglichen die Learnings aus einer Kundensegmentierung den gezielteren Einsatz finanzieller Ressourcen. Newsletter können gezielt für stark besetzte Kundensegmente entwickelt werden, ein Katalogversand auf die für Katalogwerbung empfänglichen Gruppen beschränkt werden. Darüberhinaus lässt sich Online-Werbung gezielter aussteuern: Sobald über einen Cookie die Verknüpfung zwischen dem Kundenaccount und einer aktuellen Session auf eine anderen Website hergestellt wird, kann automatisiert auf Basis des Kundensegments entschieden werden, welches Werbemittel zum Einsatz kommt. Durch die Summe der Maßnahmen sparen Sie zum einen Geld, weil Werbemaßnahmen, die für einzelne Segment irrelevant sind, gar nicht erst zum Einsatz kommen und Streuverluste minimiert werden. Zum anderen haben Sie die

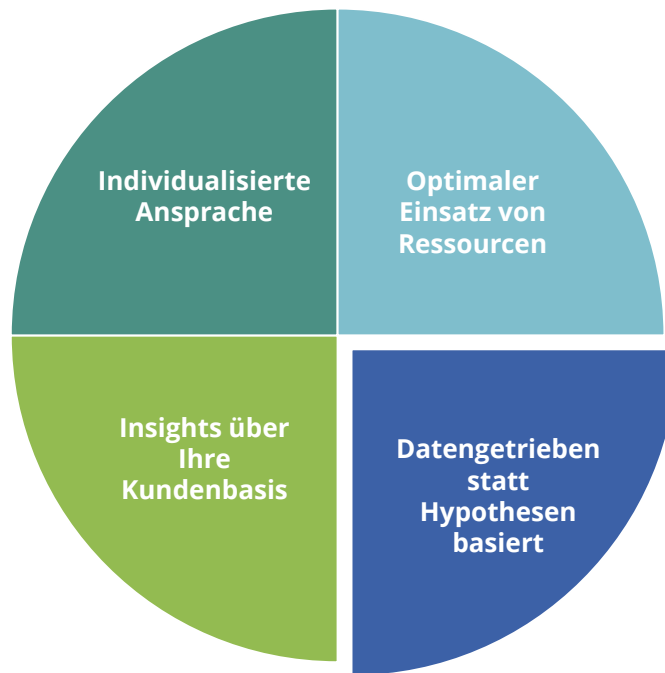


Abbildung 1: Gründe für eine datengetrieben Kundensegmentierung

Möglichkeit, maßgeschneiderte Werbebotschaften einzusetzen.

Datenbasis einer Kundensegmentierung

Für eine Kundensegmentierung sollten Sie alle Daten in Betracht ziehen, die Informationen zur umfassenden Charakterisierung Ihrer Kunden enthalten. Dies reicht von demografischen Daten über Bestellhistorie und Tracking-Daten bis hin zu von Ihnen durchgeführten Umfragen, die z.B. Kundenzufriedenheit und Freizeitaktivitäten erfassen. Ausreichend für den Beginn einer Segmentierung sind bereits Daten zur Bestellhistorie Ihrer Kunden. Informationen zur Anzahl der Bestellungen und bestellten Produkte, zu Preisen und Retouren, etc. finden sich meistens standardmäßig in Ihrem CRM-System. Auch rein auf Basis von Tracking-Daten lässt sich eine Kundensegmentierung durchführen. Ein Zusammenspiel von CRM- und Tracking-Daten ist insbesondere bei Geschäftsmodellen mit niedriger Bestellfrequenz hilfreich, um hochfrequente Informationen zu erhalten und auch potenzielle Neukunden zu berücksichtigen.

Relevante Metriken

Jegliche Informationen die für die differenzierte Betrachtung Ihrer Kunden notwendig und verfügbar sind, sollten bei der Analyse berücksichtigt werden. Um hier-

bei eine möglichst umfassende Perspektive zu erhalten, empfehlen wir ein Kick-off-Meeting mit Vertretern aus den unterschiedlichen Fachabteilungen. Der abteilungsübergreifende Austausch ist notwendig, um sicherzustellen, dass keine relevanten Informationen übersehen werden. Zusätzlich werden die auf Basis der Kundensegmente auszusteuern Business Cases bereits frühzeitig diskutiert und leisten auch im Anschluss an die Segmentierung wertvolle Dienste, wenn die Interpretation der ermittelten Segmente bevorsteht.

Metriken: Beispiele

Einige Kundenmetriken sind weitestgehend unabhängig vom Geschäftsmodell und bei nahezu jeder Kundensegmentierung im B2C-Bereich relevant, so z.B.:

- Alter
- Geschlecht
- Bestellfrequenz
- Warenkorbwerte
- Produkte
- Retourenquote
- Verwendung von Gutscheinen/Rabatten
- Zahlungsart

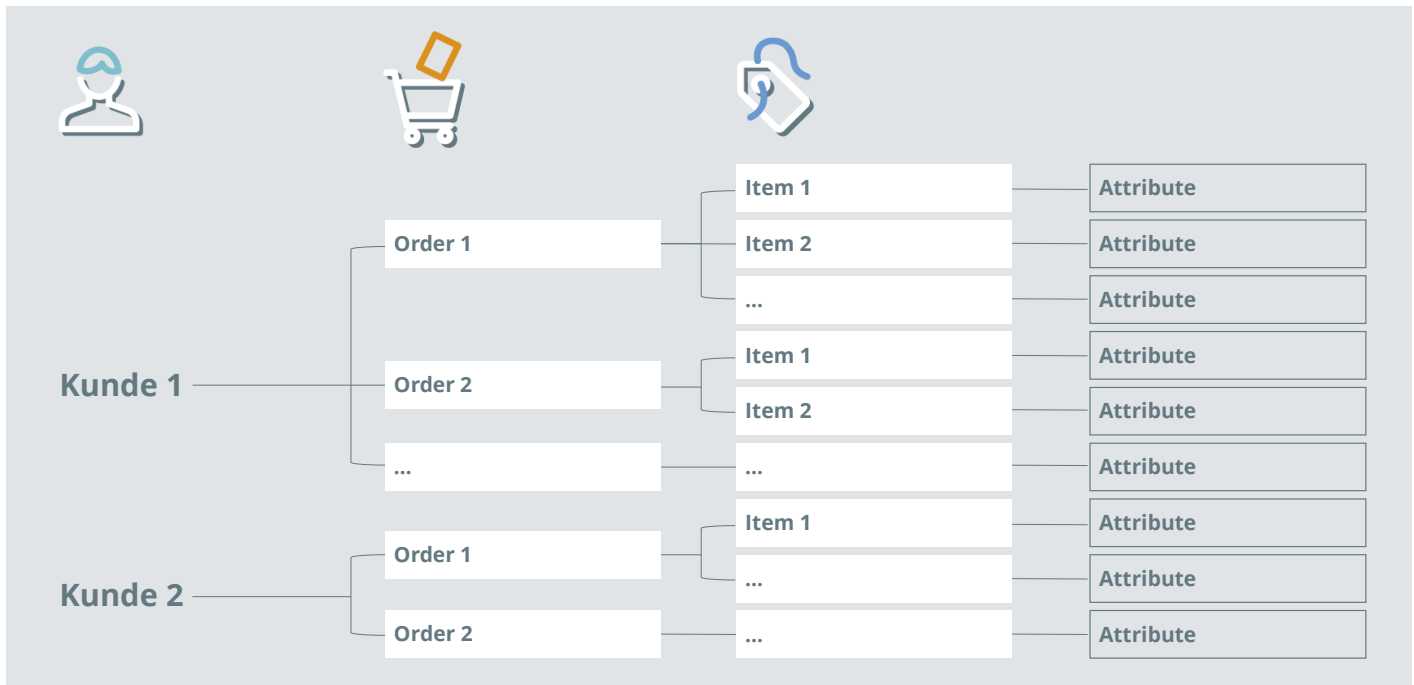


Abbildung 2: Typische Struktur von Bestelldaten: Für jeden Kunden liegen eine oder mehrere Bestellungen vor, die sich aus einem oder mehreren Items zusammensetzen. Um die Daten in der Kundensegmentierung zu nutzen, ist eine Aggregation auf Kundenebene notwendig.

Auswahl des Analysezeitraums

Um eine ausreichende Datenlage zu gewährleisten, sollte der Analysezeitraum ausreichend groß sein. Allerdings ist es nicht unbedingt notwendig, die gesamte Kundenhistorie zu berücksichtigen.

Die Länge der benötigten Datenhistorie ist in erster Linie abhängig von der typischen Bestellfrequenz der Kunden. So leuchtet es ein, dass z.B. bei der Betrachtung von Reisebuchungen ein längerer Zeitraum (1-2 Jahre) der Bestellhistorie berücksichtigt werden sollte, während bei Bestellungen von Lebensmitteln für den täglichen Gebrauch schon Wochen oder wenige Monate ausreichend sind. Mitunter können die Informationen aus weit zurückliegenden Bestellungen sogar irrelevant oder verfälschend sein, wenn sich das zugrundeliegende Geschäftsmodell oder die Zusammensetzung der Kundenbasis zwischenzeitlich stark verändert haben.

Weiterhin ist es sinnvoll, bei der Wahl des Zeitraums Saisonalitäten zu berücksichtigen. In der Regel variiert die Gesamtzahl der Bestellungen über die Woche und über das Jahr hinweg systematisch, beispielsweise im Rahmen typischer Urlaubszeiten oder des Weihnachtsgeschäfts. Um unerwünschte Einflüsse dieser Schwankungen auf die Analyse zu verhindern, empfiehlt es sich, mindestens einen kompletten Zyklus mit den Daten abzudecken.

Datenverdichtung

Zum Ermitteln der Kundensegmente sind Daten auf Kundenebene notwendig. Manche Daten liegen bereits auf Kundenebene vor, z.B. Alter, Geschlecht oder Herkunftsland. Andere Daten sind typischerweise granular und liegen auf Bestell- oder Articlebene vor. Abbildung 2 verdeutlicht die typische hierarchische Struktur von Bestelldaten. Diese müssen dann auf Kundenebene verdichtet werden. Hier sind Erfahrung und Abstimmung gefragt, um sinnvolle Aggregationslogiken zu entwickeln. Abbildung 4 zeigt exemplarisch, wie sich derartige Daten auf Kundenebene verdichten lassen.

Statistische Herangehensweise

Die sogenannte Clusteranalyse ist *das* datenbasierte Verfahren zur Identifikation von Kundensegmenten. Bei ihr handelt es sich um ein struktur-entdeckendes Verfahren, welches unbekannte Zusammenhänge zwischen Untersuchungseinheiten, in Ihrem Fall Ihren Kunden, herausarbeitet. Im Rahmen der statistischen Analyse werden üblicherweise mehrere konzeptionell unterschiedliche Methoden eingesetzt und nachgelagert die optimale Clusterlösung identifiziert.

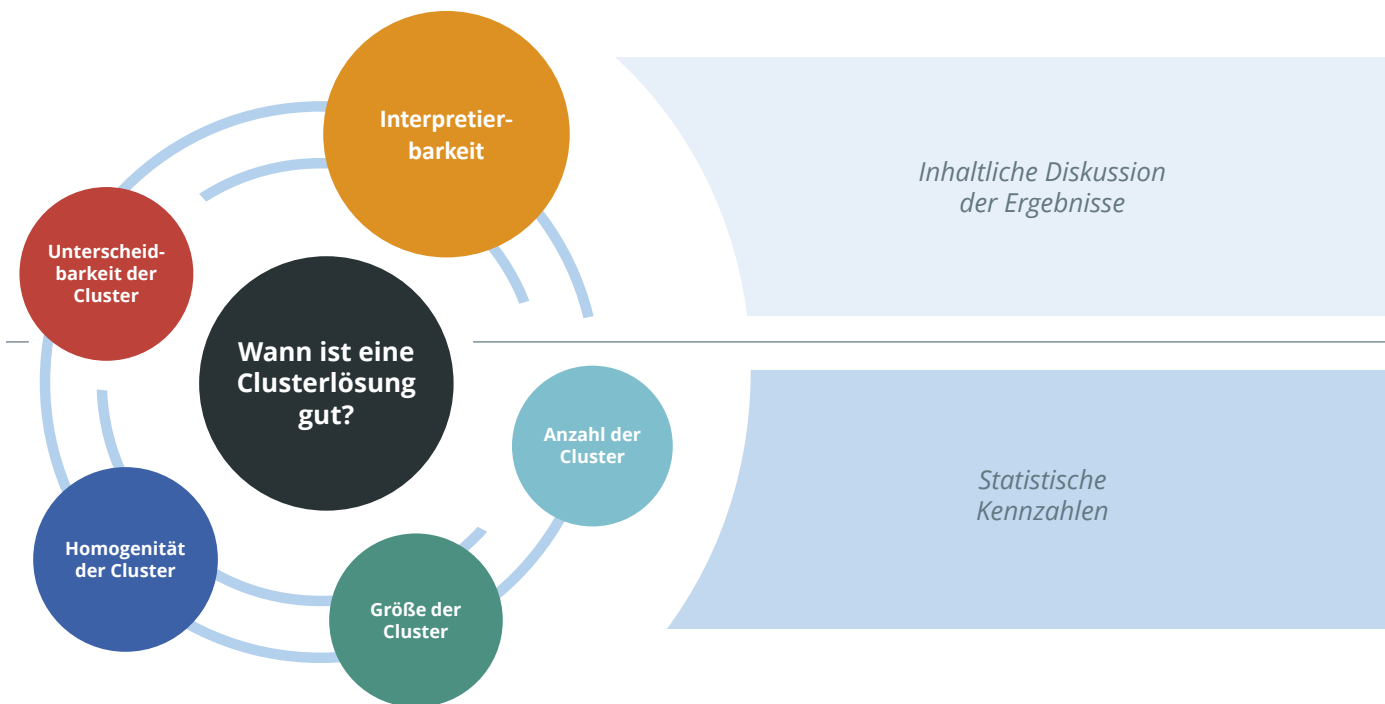


Abbildung 3: Kriterien zur Beurteilung der Güte einer Clusterlösung in der Kundensegmentierung. Am wichtigsten ist die Interpretierbarkeit, weil sie eine produktive Nutzung der Kundensegmente im Unternehmen ermöglicht.

Methoden zur Clusteranalyse

Es gibt eine Vielzahl unterschiedlicher statistischer Methoden zur Durchführung einer Clusteranalyse. "Unter der Haube" unterscheiden sie sich diese bezüglich des Algorithmus, mit dem die einzelnen Kunden in Gruppen eingeordnet werden. Generell lassen sich die Algorithmen in drei große Gruppen einteilen: *hierarchische*, *partitionierende* und *modellbasierte* Ansätze.

Hierarchische Methoden

Hierarchischen Methoden basieren auf den folgenden zwei Schritten:

1. Ermittlung eines Ähnlichkeitsmaßes zwischen den einzelnen Kunden.
2. Darauf basierend: Ermittlung der idealen Einteilung der Kunden in Cluster/Segmente, sodass die Kunden innerhalb eines Clusters sehr ähnlich sind, sich zwischen den Clustern jedoch stark unterscheiden.

Hierarchische Clustermethoden trennen sich in zwei Herangehensweisen auf: Agglomerative Verfahren beginnen auf der Ebene einzelner Kunden und fassen die einander ähnlichsten schrittweise zu immer größeren Gruppen zusammen. Divisive Verfahren beschreiten den umgekehrten Weg und zerteilen die Gesamtmenge der

Kunden in nach und nach immer kleinere Untergruppen. Die Zusammenlegung bzw. Aufteilung basiert dabei stets auf der Ähnlichkeit der Kunden(-gruppen) zueinander. In beiden Szenarien erhält man die namensgebende hierarchische Relation der Gruppen: jedes Cluster setzt sich aus jeweils zwei unterschiedlichen Untergruppen zusammen (s. Abb. 5). Dadurch können auch in der späteren Implementierung relativ unproblematisch mehrere Lösungen mit einer unterschiedlichen Anzahl an Segmenten parallel vorgehalten und je nach Business Task angesprochen haben.

Partitionierende Methoden

Bei partitionierenden Clustermethoden werden die Kunden den Zentren der jeweiligen Segmente, die sogenannte Medoide, zugeordnet, zu dem sie den geringsten Abstand aufweisen. Da die Segmente und damit deren Medoide zunächst unbekannt sind, ist eine mehrfache Berechnung mit zufällig ausgewählten ersten Zentren notwendig, um sicherzustellen, dass der iterative Algorithmus die optimale Lösung unabhängig von den zufälligen Startwerten berechnet. Die Medoide der finalen Segmente verkörpern den typischen Repräsentanten des Segments.



Abbildung 4: Der Kundensegmentierung ist der Schritt der Datenverdichtung vorgelagert.

Modelbasierte Ansätze

Für die Durchführung einer modellbasierten Segmentierung stehen eine Vielzahl von Algorithmen zur Verfügung. Allen Ansätzen ist die Idee gemein, dass die Kunden per se zu unterschiedlichen Gruppen gehören, sich diese Gruppen aber nicht direkt beobachten lassen. Die Zugehörigkeit zu einer solchen nicht beobachtbaren, latenten Gruppe ist wiederum ursächlich für das Verhalten der Kunden und manifestiert sich in den zur Verfügung stehenden Kundendaten. Im Rahmen der Modellierung werden dann die einzelnen potenziellen Segmente explizit definiert und es wird geprüft, welches Set an Definitionen die Kundendaten am besten modellieren bzw. reproduzieren kann.

Welche Methode und welche Clusterlösung sind die richtigen?

Die Frage, die sich in Anbetracht der Vielzahl an möglichen Clusterverfahren stellt, ist die nach der richtigen Methode. Die Antwort auf diese Frage ist beides, einfach und schwierig zugleich. Die einfache, aber sehr allgemeine Antwort lautet: das Verfahren ist das richtige, das es schafft aus den zur Verfügung stehenden Daten eine *optimale* Clusterlösung zu ermitteln. Diese sehr allgemeine Antwort ist darin begründet, dass es sich bei der Clusteranalyse wie bereits angemerkt um ein struktorentdeckendes und nicht um ein strukturprüfendes Verfahren handelt. Deshalb kann die Güte nicht direkt anhand einer klar definierten Größe, wie z.B. der Prognosequalität, beurteilt werden kann. Die indirekte Bewertung basiert deshalb auf der folgenden Frage:

Wann ist eine Clusterlösung optimal?

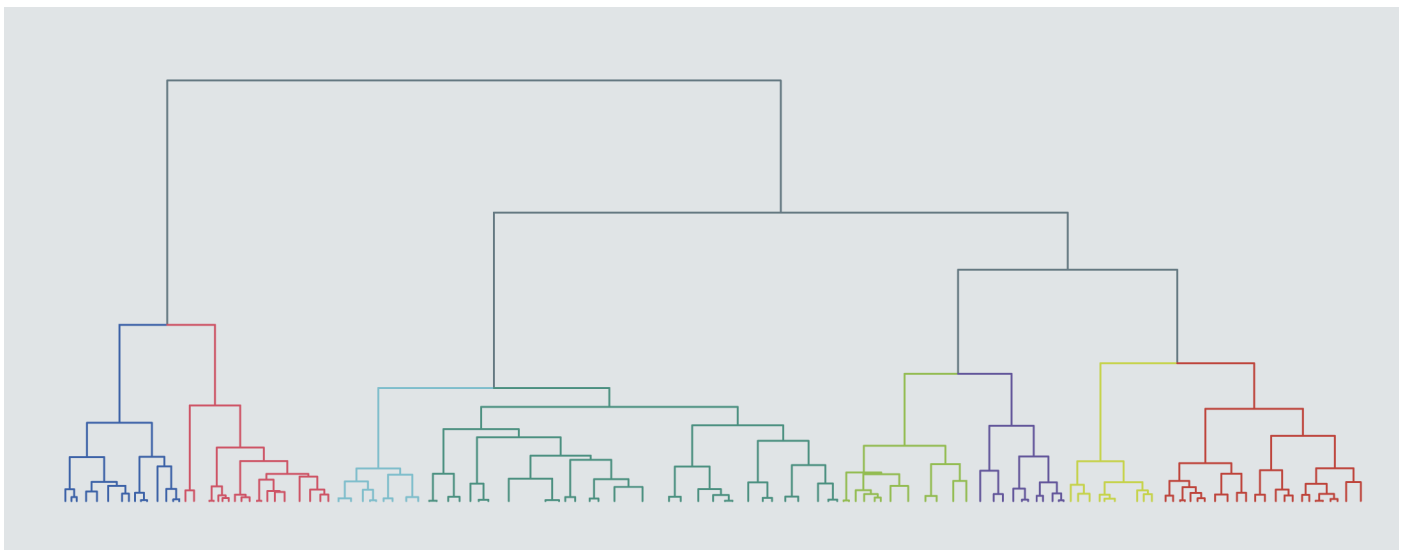
Diese Beantwortung dieser Frage dient nicht nur der Festlegung auf das Clusterverfahren, sondern auch dazu die a priori ebenfalls unbekannt Anzahl der Segmente zu ermitteln. Die Kriterien die zur Beurteilungen herangezogen werden sind in Abbildung 3 dargestellt.

Statistische Kennzahlen

Kriterien wie Homogenität innerhalb der Segmente, sowie Größe und Anzahl der Segmente lassen sich einfach berechnen. Die Kriterien Homogenität, Anzahl und Größe sind jedoch nicht unabhängig voneinander, sondern weisen gegenläufige Tendenzen auf. Deshalb werden für die Ermittlung des *statistischen* Optimums Kennzahlen herangezogen, die in Abhängigkeit der Clusteranzahl die Kompaktheit und den Abstand zwischen den Segmenten berücksichtigen. Betrachtet man diese in Abhängigkeit der Segmentanzahl und für unterschiedliche Methoden, so lassen sich aus statistischer Sicht vielversprechende potenzielle Lösungen identifizieren.

Inhaltliche Diskussion

Das A und O einer guten Clusterlösung ist ihre Interpretierbarkeit. Ergibt sich beim Betrachten der Eigenschaften eines Clusters ein klares Bild der entsprechenden Kundengruppe bzw. der die Gruppe repräsentierenden Persona, so ist dies die beste Voraussetzung dafür, Geschäftsprozesse maßgeschneidert auf diese Gruppe anzupassen. Erscheinen die Cluster hingegen diffus und unklar, sollte geprüft werden, ob andere Lösungen bes-



Kunden-ID

Abbildung 5: Schematische Darstellung des hierarchischen Clustering. Es ist gut erkennbar, wie in jedem Schritt zwei Cluster zu einem zusammengefasst werden. Je weniger Cluster, desto größer wird die Heterogenität der Segmente.

sere Ergebnisse liefern. Ein weiterer Diskussionspunkt ist die Handhabbarkeit der Anzahl und Größe der Segmente. So kann eine Lösung mit vielen und gut interpretierbaren Segmenten für die Anwendung nicht geeignet sein, da die Zahl der Segmente die Kapazitäten für eine individualisierte Ansprache übersteigt oder zu geringe Fallzahlen eine Individualisierung unrentabel werden lassen. Die Berücksichtigung aller angesprochenen Aspekte ist also notwendig, um eine Clusterlösung zu beurteilen. Hervorzuheben ist, dass dieser Prozess im Rahmen der Analyse durch ein umfangreiches Reporting zu den Segmentcharakteristika vorbereitet wird, die eigentliche Diskussion aber mit allen Stakeholdern gemeinsam geführt werden sollte, um die für Ihr Unternehmen optimale Lösung zu identifizieren.

Ergebnisse einer Kundensegmentierung

Nach Abschluss der Analyse und der Diskussion der Segmente, verfügen Sie über die Information, aus welchem Umfang von Personas sich Ihre Kunden zusammensetzen. Sie können durch diesen datengetriebenen Ansatz zum Verständnis Ihrer Kundenbasis das Arbeiten mit hypothetischen Personas ablösen. Ein Abgleich mit den postulierten Hypothesen, versetzt Sie in die Lage zu ermitteln, welche Kundengruppen in Relation zu Ihren Annahmen über- oder unterrepräsentiert sind. Darüberhinaus haben Sie für jeden bei der Analyse berücksichtigten Kunden direkt die Einordnung in das entsprechende Segment. Dadurch können Sie sofort mit der Umsetzung

segmentspezifischer Aktionen starten und deren Nutzen z.B. im Rahmen eines A/B Tests quantifizieren. Zusätzlich liefert die Analyse einen Algorithmus, der es Ihnen erlaubt weitere Kunden und Kunden mit verändertem Nutzungsverhalten den identifizierten Segmenten zuzuordnen. Es empfiehlt sich diesen Algorithmus in Ihrem BI-System zu implementieren und somit sicherzustellen, dass automatisch aktuelle Segmentzuordnungen verfügbar sind. Sollten für den Live-Betrieb nicht alle Metriken zur Verfügung stehen, die zur Analyse herangezogen wurden, wenn diese z.B. aus einem zusätzlichen Survey stammen, bedeutet das nicht, dass für Sie eine Ausweitung der Ergebnisse auf neue Kunden unmöglich ist. Durch einen zusätzlichen Klassifikationsalgorithmus bietet sich die Möglichkeit die Learnings aus dem Segmentierungsprojekt auf eine neue, schmalere Datenbasis zu übertragen.

Fazit

Wir erarbeiten in engem Austausch mit Ihnen eine maßgeschneiderte Lösung, die datengetrieben Ihre tatsächlich vorhandenen Kundensegmente ermittelt. Die Kombination aus Ihrem Domänenwissen und unserer Data Science-Expertise stellt sicher, dass die identifizierten Segmente für Sie direkt als Personas greifbar werden. Die Ergebnisse der Analyse können Sie unmittelbar nutzen, um Ihre Geschäftsprozesse segment-spezifisch zu optimieren. Eine auf die Segmentierung aufbauende automatisierte Einordnung von Neukunden wird ebenfalls möglich.

Ihr Ansprechpartner





Dr. Steffen Wagner

Steffen ist Mitgründer von INWT. Er beschäftigt sich schwerpunktmäßig mit den Themen Predictive Analytics, Online Marketing und Customer Relationship Management. Der promovierte Physiker gibt als Dozent im 'Joint Masters Program of Statistics' in Berlin Einblicke in seine Tätigkeit als Data Scientist.

Kontakt

- **Tel.:** +49 (0)30 1208231-0
- **E-Mail:** steffen.wagner@inwt-statistics.de

 **INWT Statistics GmbH**
Hauptstraße 8
Meisenbach Höfe, Aufgang 3a
10827 Berlin

 +49 30 1208231-0

 info@inwt-statistics.de

 www.inwt-statistics.de

