

# DER NUTZEN VON TRANSAKTIONS DATEN FÜR DAS HANDELSMARKETING

Hendrik Schröder, Andreas Rödl



aus: Trommsdorff, V. (Hrsg.)

## Handelsforschung 2004

BBE-Verlag, Köln

S. 519-538

# DER NUTZEN VON TRANSAKTIONS DATEN FÜR DAS HANDELSMARKETING

Hendrik Schröder, Andreas Rödl

„Wir dürsten nach Informationen, aber wir ertrinken in Daten!“ Diesen Hilferuf hört man allzu häufig im Einzelhandel. Tagtäglich liefern ihm die Kunden eine Fülle von Daten, wenn sie bei ihm einkaufen. Diese Transaktionsdaten können anonym, pseudonym oder personalisiert sein. Je mehr der Kunde aus der Anonymität heraustritt, desto mehr Möglichkeiten ergeben sich bei der Datenanalyse und bei der Gestaltung von Marktbearbeitungsmaßnahmen. Der folgende Beitrag charakterisiert zunächst die verschiedenen Arten von Transaktionsdaten und zeigt anschließend auf, welche für das Handelsmarketing relevanten Fragen sich mit ihnen beantworten lassen.

## 1. EINLEITUNG

Jeden Tag fallen im Einzelhandel bei den Transaktionen mit den Kunden Millionen von Daten an. Sie liefern Auskunft über Art und Anzahl der gekauften Produkte, Ort und Zeitpunkt des Kaufes, die Nutzung von Rabatten sowie andere Sachverhalte. Teils sind es anonyme Daten, die die Identität des Kunden nicht preisgeben, wie z.B. Scannerdaten, teils werden die Umsatzdaten mit persönlichen Daten der Kunden verknüpft, wie z.B. beim Einsatz von Kundenkarten im stationären Einzelhandel oder bei der Abwicklung von Käufen im Distanzhandel (Katalogversand, Electronic Retailing etc.).

Die Technik ist mittlerweile so weit fortgeschritten, dass die Verarbeitung und Analyse der Daten kaum noch den Engpass bildet. Gleichwohl muss man feststellen, dass viele Handelsunternehmungen vorhandene Daten nur in geringem Maße nutzen, um die Planung und Kontrolle von Marketingmaßnahmen zu unterstützen. Wenn Daten verarbeitet und analysiert werden, handelt es sich meist eher um Vorgänge niedriger Komplexität, wie z.B. Renner-Penner-Listen. Die Zurückhaltung, vorhandene Daten zu nutzen, ist vor allem aus Problemen bei den „Fragestellungen“, „Datenquellen“ und „Methoden“ zu erklären: Der Anwender muss wissen, welche Fragestellungen des Handelsmarketing er mit Daten aus welchen Quellen und mit Hilfe welcher Methoden beantworten kann. Herkömmlicher Sicht nach ist die Fragestellung bekannt und man sucht nach geeigneten Quellen und Methoden, um den Informationsbedarf zu decken (Verwendungsorientierung, Anlassorientierung). Der Reiz neuer Techniken liegt dagegen in der Möglichkeit

aufzuzeigen, welche Fragen sich mit Hilfe vorhandener Daten und Methoden beantworten lassen (vgl. Knobloch 2000, S. 9 f.). Diese Sichtweise (Potenzialorientierung, Datenmustererkennung) greift der vorliegende Beitrag auf. Die folgenden Ausführungen wollen auch dazu beitragen, dem Customer Relationship Management wieder mehr Optimismus entgegenzubringen, der durch zahlreiche Projekte mit falschem, weil einseitig auf die Technik ausgerichteten Verständnis für CRM und fehlendem Verständnis für die relevanten Fragen der Anwender vielerorts auf der Strecke geblieben ist (vgl. Schröder, Feller, Schuck 2002).

## **2. ARTEN VON TRANSAKTIONSDATEN**

### **2.1 Zur Systematisierung von Transaktionsdaten**

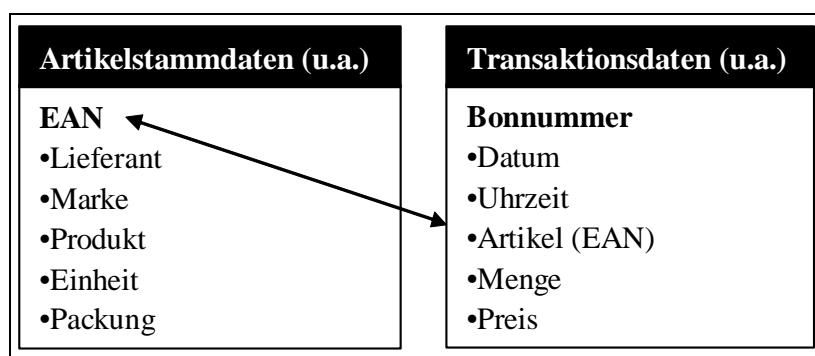
Daten, die im Zusammenhang mit Kaufvorgängen anfallen, lassen sich nach verschiedenen Merkmalen systematisieren. Merkmale können sein der *Ort der Erhebung* (innerhalb oder außerhalb der Einkaufsstätte), die *Institution der Erhebung* (die Handelsunternehmung selbst oder Dritte, wie z.B. Marktforschungsinstitute), der *Verbund von Datenquellen* (reine Abverkaufsdaten, Verbindung von Abverkaufsdaten mit Artikelstammdaten, Verbindung von Abverkaufsdaten mit Kundendaten) und der Grad der *Personalisierung* (anonym, pseudonym oder personalisiert).

Transaktionsdaten eines Kunden, die mit *Scannerkassen* erfasst werden, fallen folglich intern an, werden von der Handelsunternehmung im Zuge des Kaufvorganges selbst erhoben, lassen sich mit weiteren Datenarten, wie z.B. Artikelstammdaten verbinden, liefern aber ohne kundenindividuelle Daten nur anonyme Ergebnisse. Transaktionsdaten, die ein Marktforschungsinstitut im Rahmen eines *Verbraucherpanels* erhebt, können außerhalb der Einkaufsstätte ermittelt werden, z.B. durch das Inhome-Scanning, lassen sich konkreten Personen oder Haushalten zuordnen und sind pseudonym, d.h. die Identität der Personen oder Haushalte bleibt gewahrt. Transaktionsdaten, die unter Verwendung einer *Kundenkarte* oder bei der Abwicklung von Käufen im *Distanzhandel* anfallen, haben immer das Potenzial verbundener Daten, da sich die Transaktionsdaten mit kundenindividuellen Daten (Name, Adresse, Alter etc.) anreichern lassen, und bieten das Potenzial personalisierter Daten, da jeder Kunde identifiziert und individuell angesprochen werden kann. Welche Eigenschaften anonyme, pseudonyme und personalisierte Transaktionsdaten aufweisen, erläutern die folgenden Ausführungen.

## 2.2 Anonyme Transaktionsdaten

Die wichtigste Quelle zur Gewinnung von Transaktionsdaten im stationären Einzelhandel sind *Scannerkassen*. Fast alle großen Unternehmungen des Lebensmitteleinzelhandels (Metro, Edeka, Spar, Markant, Tengelmann, dm, Schlecker) sind in 80 bis 100% ihrer Betriebe mit solchen Kassen ausgestattet. Lediglich Rewe fällt mit einem Ausstattungsgrad von weniger als 50% aus dem Rahmen (vgl. Schäfer 2002).

Wenn beim Kassivorgang der Preis einer Ware durch Scannen des Barcodes ermittelt wird und nicht mehr durch die manuelle Eingabe des Preises in eine „Rechenmaschine“, enthält ein Bon neben Preisinformationen auch die zugehörige Artikelidentifikationsnummer (Internationale Artikel Nummer – EAN). Um die vom Kunden zu zahlende Gesamtsumme zu ermitteln, werden die Preise aller gekauften Artikel addiert und gleichzeitig die EAN einer Bonnummer zugeordnet. Über die EAN können die in den Artikelstammdaten des Warenwirtschaftssystems gespeicherten Informationen mit den Transaktionsdaten verknüpft werden (vgl. *Abbildung 1*).



**Abbildung 1: Datenmodell von Transaktionsdaten in einem Scannerkassensystem**

Durch den Zwang der Preisermittlung wird gewährleistet, dass alle Bons vollständig erfasst werden. Allerdings schöpfen nicht alle Scannerkassensysteme ihr Potenzial aus. Mancherorts beschränken sie sich auf die Funktion der effizienten Preisermittlung in der Check-Out Zone; die Speicherung und Analyse von Bons werden dagegen vernachlässigt. Schwerwiegender sind die Probleme, die sich beim Scanning ergeben können (vgl. hierzu ausführlich Ebberts 2003, S. 148 f.), wie etwa das Fehlscanning. Dies tritt zum Beispiel auf, wenn Joghurt derselben Marke, den ein Kunde in mehreren Geschmacksrichtungen (z.B. Erdbeere, Banane, Schoko) kauft, vom Kassenpersonal nur einmal gescannt (z.B. nur Erdbeere) und somit nur eine Geschmacksrichtung als mehrfacher Kauf im Umfang aller Joghurtbecher dieser Marke gemessen wird. Ein weiteres Problem ist

die so genannte „Sumpftaste“, über die man alles erfasst, was sich nicht gut scannen lässt oder wo versäumt wurde, neue EAN oder Preise in das Kassensystem einzuspielen. Produkte, die nicht oder fehlerhaft gescannt werden, verfälschen die Datenbasis. Weitere Probleme ergeben sich aus fehlerhaften Datenbeständen im Warenwirtschaftssystem und aus der Verwendung proprietärer Artikelnummernsysteme. Dies erschwert die Verknüpfung mit externen und anderen internen Daten und verhindert insbesondere die eindeutige Zuordnung von Artikelinformationen.

Scannerdaten sind auch die Grundlage für *Handelspanels*. Marktforschungsinstitute erhalten von verschiedenen Handelsbetrieben Scannerdaten, die sie auswerten und als Ergebnisse den Händlern und Herstellern zur Verfügung stellen. Das Prinzip der Datenerhebung ist identisch mit der Erhebung eigener Scannerdaten. Die Teilnahme mehrerer Handelsbetriebe am Panel erlaubt bei der Auswertung zeitpunkt- und zeitraumbezogene Betriebsvergleiche, also Quer- und Längsschnittanalysen. Zu beachten ist, dass sich die bei *einem* Handelsbetrieb auftretenden Probleme des Scanning multiplizieren, was die Reliabilität der Daten weiter einschränkt. Für viele Hersteller sind Handelspanels allerdings die einzige Möglichkeit, Abverkaufsdaten der Händler zu erhalten.

### **2.3 Pseudonyme Transaktionsdaten**

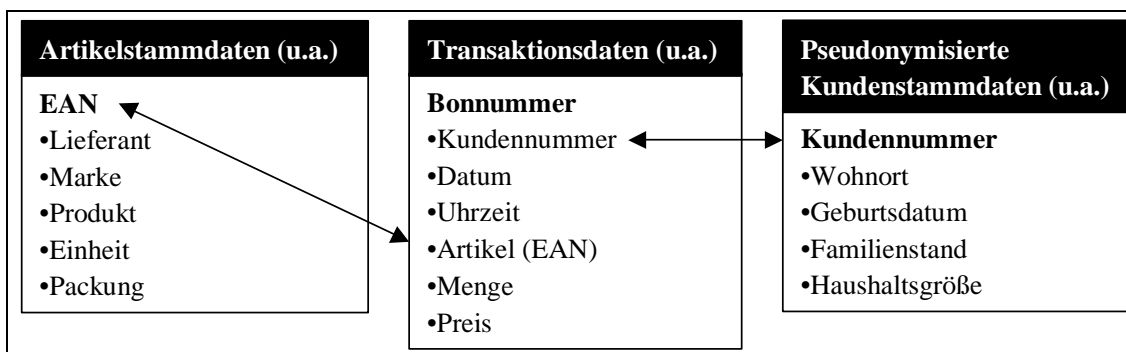
Pseudonyme Transaktionsdaten erlauben die Zuordnung von Käufen zu bestimmten Kunden oder Haushalten, ohne deren Namen und Adresse zu verwenden. Anstelle des Namens findet sich ein Pseudonym, in der Regel eine Kundennummer oder eine Kartennummer. Soweit die unter einem Pseudonym erfassten Nutzungsprofile nicht mit Daten über den Träger des Pseudonyms zusammengeführt werden (dürfen), bleiben die Personen anonym.

Eine Quelle pseudonymer Transaktionsdaten sind *Verbraucherpanels* (Haushalts- und Individualpanels). Die Daten werden über Scanning im Einzelhandel oder durch In-Home-Scanning der Panelteilnehmer erhoben. Ein Haushaltspanel soll das Kaufverhalten des Haushaltes möglichst umfassend abbilden. Jeder aufgezeichnete Einkauf lässt sich über das Pseudonym jedem Haushalt exakt zuordnen.

Wie bei der Gewinnung anonymer Daten kann es auch hier zu Fehlern bei der Erhebung und Speicherung von Daten kommen. Wenn die Artikel nicht im Zuge des POS-Scanning (mit ID-Karte des Haushaltes), sondern im Zuge des Inhome-Scanning erhoben werden, besteht die Gefahr, dass bestimmte Artikel nicht gescannt, möglicherweise

komplette Einkäufe und damit Bons unberücksichtigt bleiben. Die geringe Fallzahl von Haushaltspanels (GfK: ca. 12.000 Haushalte, Nielsen: ca. 8.000 Haushalte) kann sich negativ auf die Aussagekraft auswirken, sobald detailliertere Analysen (Drill down) gewünscht sind, z.B. Käufe in bestimmten Geschäften. Auf weitere Probleme, wie z.B. Paneleffekte und Panelsterblichkeit, können wir an dieser Stelle nicht eingehen (vgl. dazu Meffert 1992, S. 215 f.).

Abbildung 2 verdeutlicht das Prinzip der Zuordnung verschiedener Datenquellen unter Einbeziehung von Daten aus einem Haushaltspanel. In den Transaktionsdaten selbst sind nie alle zugrunde liegenden Stammdaten (der Artikel und der Kunden) enthalten, es ist immer ein Schlüssel notwendig. Vor diesem Hintergrund ist auf folgendes Problem hinzuweisen: Bei fehlenden Zuordnungsmechanismen misslingt die Integration der Paneldaten in ein bestehendes Data-Warehouse. Dies ist z.B. dann der Fall, wenn das Marktforschungsinstitut alle Produkte mit einem eigenen Schlüssel versieht und sie ihren Kunden so zur Verfügung stellt, anstatt die EAN als Schlüssel zu verwenden. Das so genannte Matching mit unternehmensinternen Daten oder Daten anderer Anbieter ist dann nur nach einem enormen Datenbereinigungsaufwand möglich.



**Abbildung 2: Datenmodell von Transaktionsdaten in einem Haushaltspanel**

Zwei weitere Möglichkeiten, pseudonymisierte Abverkaufdaten zu erhalten, ergeben sich durch die Identifikation des Kunden über EC- oder Kreditkartennummern und über die Teilnahme an einem Kundenkartenverbund. *Bargeldlose Zahlungsmittel* weisen als Identifikationsmethode einige Schwächen auf. Die Erhebung, Verarbeitung und Nutzung personenbezogener Daten in Verbindung mit Abverkaufdaten ist ohne die Einwilligung des Kunden nicht zulässig (§ 28 Abs. 4 BDSG). Allenfalls kann von der Pseudonymisierung Gebrauch gemacht werden. Sofern ein Kunde unterschiedliche Karten nutzt, liegen jedoch verschiedene Pseudonyme für ihn vor. Weitere Probleme sind, dass viele Kunden im stationären Einzelhandel bar bezahlen und dass die Gültigkeit jeder

EC- oder Kreditkarte auf wenige Jahre begrenzt ist. Wenn nach der Ausgabe neuer Karten die Pseudonyme (Kartennummern) wechseln, sind der Längsschnittanalyse zeitliche Grenzen gesetzt.

Kundenkartenprogramme liefern pseudonyme Daten, wenn Handelsunternehmungen auf ein proprietäres Kundenkartenprogramm verzichten und sich stattdessen einem *Kundenkartenverbund* (z.B. Payback, Happy Digits) anschließen. Das Pseudonym ist die Kartenummer. Soweit sich allerdings die teilnehmenden Kunden bei der Handelsunternehmung selbst für das Kartenprogramm angemeldet haben (z.B. real oder Obi), kann der Händler alle pseudonymen Daten über die Kartenummer den Namen und Adressdaten zuordnen. Er erhält dadurch personalisierte Daten.

## 2.4 Personalisierte Transaktionsdaten

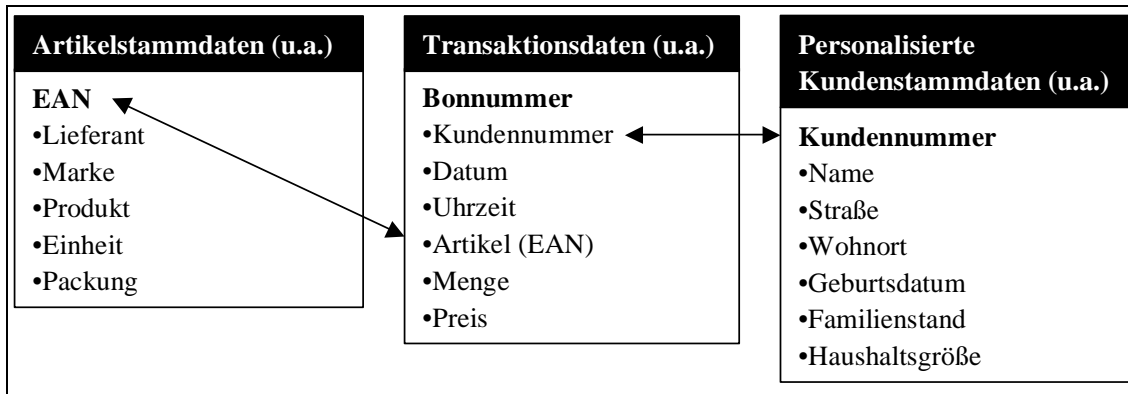
Lassen sich die Stammdaten (Name, Adresse etc.) eines Kunden mit seinen Transaktionsdaten verbinden, so liegen personalisierte Transaktionsdaten vor. Nennenswerte Quellen, personalisierte Bondaten zu generieren, sind Kundenkartenprogramme im stationären Handel und Bestellvorgänge im Distanzhandel. Die Bestelldaten im *Distanzgeschäft* liegen „zwangsläufig“ personalisiert vor, da der Kunde seine Lieferadresse angeben muss. Alle Käufe eines Kunden bei dieser Unternehmung werden im Zeitablauf also vollständig abgebildet, was entsprechende Längsschnittanalysen erlaubt. Probleme in der Zuordnung treten jedoch dann auf, wenn ein Kunde bei mehreren Käufen abweichende Lieferadressen angibt oder ihm bei mehreren Käufen abweichende Kundennummern zugewiesen werden.

Etwas anders verhält es sich bei den Daten eines Kundenkartensystems im *stationären Einzelhandel*. Der Kunde unterliegt zum einen nicht dem Zwang, sich überhaupt eine oder mehrere Kundenkarten zuzulegen, zum anderen kann er Käufe tätigen, ohne seine Karte einzusetzen. Die Konsequenz ist, dass nicht die Käufe aller Kunden als personalisierte Transaktionsdaten erfasst werden und dass Kunden mit Kundenkarten unvollständige Daten ihrer Käufe liefern.

Rechtliche Grenzen der Nutzung personalisierter Daten sind z.B. das Widerspruchsrecht der Verbraucher in § 28 Abs. 4 BDSG und die ausdrückliche Einwilligung der Betroffenen in § 5 Abs. 2 TDDSG.

Technisch funktioniert die Datensammlung, wie es von den bereits vorgestellten Methoden bekannt ist. Ein Scanningsystem registriert den Abverkauf der Ware; zusätzlich

wird nun die Kundenkarte gescannt. Somit ist der Bon um eine Position (i.d.R. eine Kundennummer) angereichert, durch die eine Verknüpfung mit der Kundenstammdatenbank möglich ist (vgl. *Abbildung 3*).



**Abbildung 3: Datenmodell von Transaktionsdaten mit einem Kundenkartensystem**

## 2.5 Vergleich verschiedener Datenquellen

Ein zentrales Kriterium zur Beurteilung von Transaktionsdaten ist der bereits zur Systematisierung herangezogene Grad der *Zurechenbarkeit* zu einzelnen Kunden. Als alleiniges Beurteilungskriterium ist die Zurechenbarkeit jedoch nicht geeignet, da je nach Zwecksetzung andere Kriterien, wie z.B. die quantitative, zeitliche und demographische Abdeckung, relevant sein können. *Abbildung 4* stellt die einzelnen Datenquellen vergleichend gegenüber.

Ein umfassendes Bild aller Käufe der eigenen Einkaufsstätte lässt sich im stationären Einzelhandel nur mit Hilfe von *Scannerdaten* gewinnen. Da diese Daten jedoch vollkommen anonym vorliegen, erlauben sie lediglich sortimentsbezogene, aber keine kundenbezogenen Längsschnittanalysen. Die Möglichkeit des Handels, die Kundenkenntnis zu erhöhen und mehr kundenorientierte Angebote zu gestalten, ist daher begrenzt.



	Anonyme Daten		Pseudonyme Daten		Personalisierte Daten	
Daten- quelle	Scanner- kasse	Handels- panel	Haushalts- panel	Kundenkar- tenverbund	Kunden- karte	Distanz- handel
Quantitative Abdeckung	nur eigene Einkaufsstät- te (EKS)	fast alle EKS	fast alle EKS	verschiedene EKS ver- schiedener Branchen	eigene EKS	eigene EKS
Zeitliche Abdeckung	eher langer Zeithorizont (seit Einfüh- rung Scan- ner-kassen)	eher langer Zeithorizont	langer Zeit- horizont (Problem: Panel- sterblichkeit)	kurzer Zeithorizont (3-6 Jahre)	kurzer Zeit- horizont (3-6 Jahre)	langer Zeit- horizont (seit Beginn d. elektron. Speicherung)
Demographische Abdeckung	Einzelbons von Kunden eines Händ- lers	Einzelbons von Kunden mehrerer Händler	Datenbasis: ca. 8.000 - 20.000 Haushalte	bis zu 22 Mio Kunden in Deutsch- land (Pay- back)	abhängig von Zahl ausgegebe- ner Kunden- karten	alle Kunden eines Händ- lers
Abbildung des Kaufverhaltens	nur ein Kauf	nur ein Kauf	Erfassung aller Einkäu- fe des Kun- den	abhängig von Einsatz und Verbrei- tung der Karte	abhängig von Einsatz und Verbrei- tung der Karte	Erfassung aller Käufe jedes Kun- den bei ei- nem Händler
Aggregation	detaillierte Informatio- nen über Rahmen- daten	Durchschn.- info (Preis, Promotions etc.), häufig auf SKU- Ebene	Aggregation zentraler Informatio- nen (Preis, Promotions etc.)	verschiedene Aggregati- onsebenen (SKU, Wa- rengruppe)	detaillierte Informatio- nen über Rahmen- daten	detaillierte Informatio- nen über Rahmen- daten
Datenqualität	hoher Berei- nigungs- aufwand notwendig	durch MaFo- Institut auf- bereitet	durch MaFo- Institut auf- bereitet	Stammdaten- abgleich ver- schiedener Händler pro- blematisch	hoher Berei- nigungs- aufwand notwendig	hoher Berei- nigungs- aufwand notwendig

**Abbildung 4: Vergleich verschiedener Quellen für Transaktionsdaten**

Einen Schritt weiter gehen *Kundenkartendaten*. Sie erlauben die dauerhafte Beobachtung bestimmter Kunden und geben im Rahmen von Längsschnittanalysen Aufschluss über das Kundenverhalten in der Einkaufsstätte. Als Begrenzung wirkt, dass nicht alle Kunden eine Kundenkarte besitzen (wollen) und dass nicht alle Kundenkarteninhaber diese auch immer einsetzen. Auch führt das Ergebnis, dass Kundenkarteninhaber höhere Bons haben als Nicht-Kundenkarteninhaber, nicht selten zu dem Trugschluss, dass die Kundenkarte zur Kundenbindung beigetragen habe. Dies ist aber unzutreffend, wenn sich gerade so genannte Heavy Buyers entschlossen haben, die Kundenkarte zu akzeptieren, oder wenn die Karte eher bei größeren Anschaffungen eingesetzt wird.

*Haushaltspanels* liefern pseudonyme Daten und gestatten kundenbezogene Längsschnittanalysen nicht nur auf Einkaufsstättenebene, sondern auch – und das ist das Alleinstellungsmerkmal dieser Datenquelle – auf Produktebene. Denn sämtliche Einkäufe eines Kunden in sämtlichen Einkaufsstätten werden erfasst. Der Hersteller erfährt von allen Panelteilnehmern, wie oft sie sein Produkt gekauft oder auch nicht gekauft haben. Jeder Handelsbetrieb muss aber davon ausgehen, dass nur ein kleiner Teil der Stichprobe auf ihn entfällt, er also wenige Aussagen über seine Abverkäufe erhält (Drill down-Problem). Daher ist es sinnvoll, sich einem Kundenkartenprogramm anzuschließen und darüber pseudonyme oder personalisierte Transaktionsdaten zu erhalten.

Abbildung 5 verdeutlicht, welchen Umfang an Käufen die verschiedenen Datenquellen abdecken. So erfasst z.B. die Einkaufsstätte 4 mit einer Scannerkasse alle Einkäufe in dieser Einkaufsstätte, das Haushaltspanel aber nur einen Teil davon.

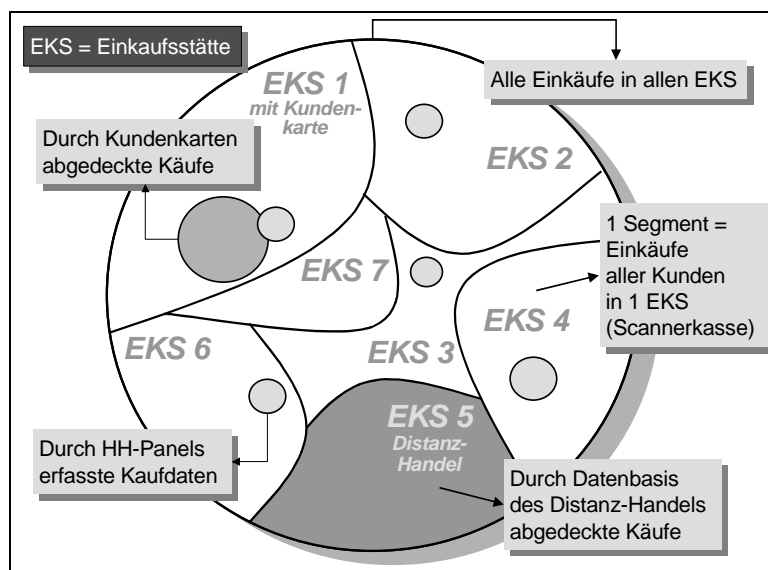


Abbildung 5: Quantitative Abdeckung der Käufe einer Einkaufsstätte

### **3. ANALYSEMETHODEN UND EINSATZGEBIETE**

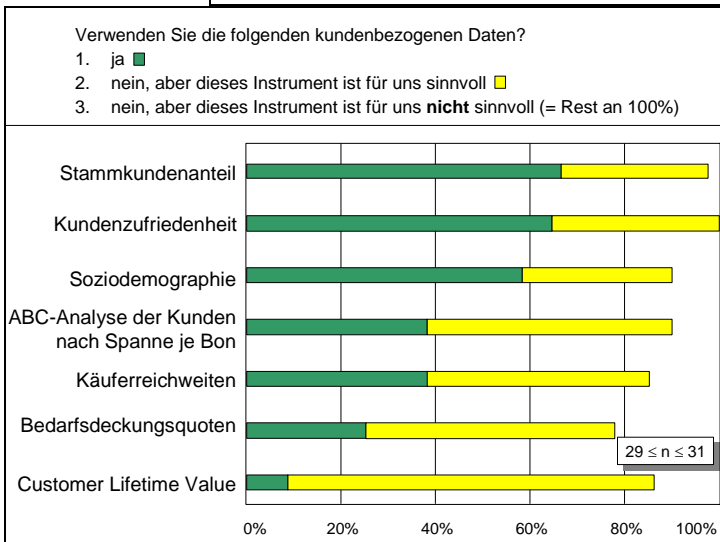
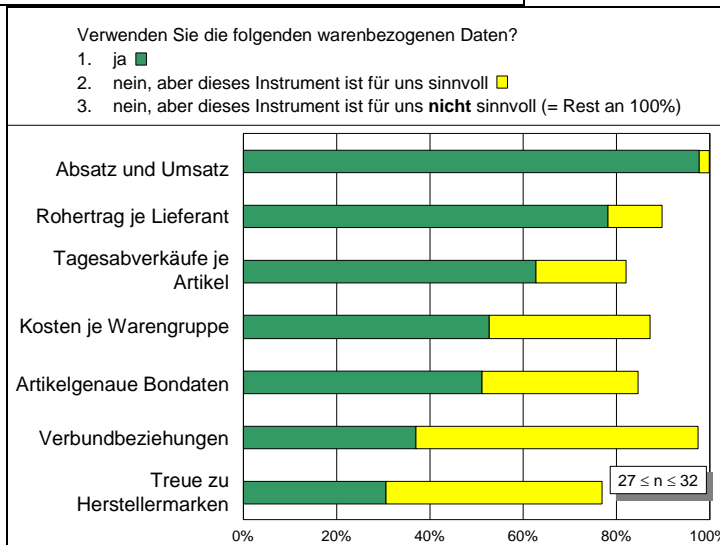
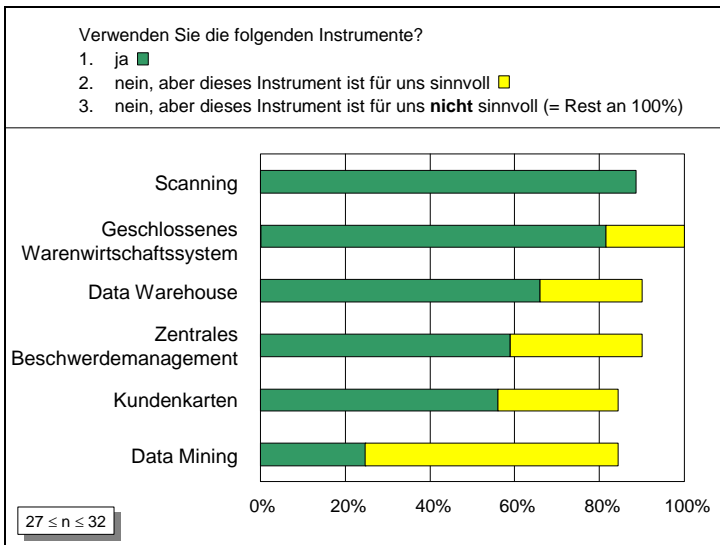
#### **3.1 Status quo der Informationsversorgung im deutschen Handel**

*Abbildung 6* vermittelt einen Eindruck, in welchem Umfang deutsche Handelsunternehmen sich mit der Erhebung und Auswertung von Daten befassen. An der im Jahr 2002 durchgeführten Studie „Consumer Insights – Was weiß der Handel über seine Ware und seine Kunden?“ des Lehrstuhls für Marketing & Handel der Universität Essen haben 32 Handelsunternehmen teilgenommen, die einen wesentlichen Teil des Umsatzes im deutschen Einzelhandel abdecken. Während die Ausstattung mit Scannerkassen bei über 80% liegt und Kundenkarten in knapp 60% der antwortenden Unternehmen eingesetzt werden, verfügen nur etwas mehr als 20% dieser Handelsunternehmen über Instrumente des Data Mining.

Betrachtet man weiterhin in *Abbildung 6*, welche Daten die Händler über die Ware und die Kunden besitzen, so zeigt sich, dass einfache Daten dominieren (Absatz, Umsatz, Stammkundenanteil etc.). Komplexere und detailliertere Informationen wie Verbundbeziehungen und Customer Lifetime Value verwenden erst wenige Unternehmen. Angesichts des geringen Verbreitungsgrades von Data Mining kann es nicht verwundern, dass damit gewonnene Informationen ebenfalls nicht weit verbreitet sind. Auch ist erkennbar, dass kundenbezogene Informationen in geringerem Ausmaß verfügbar sind als warenbezogene Informationen. Dies überrascht nicht, denn der Ausstattungsgrad der Handelsunternehmen mit Kundenkarten beträgt zwar knapp 60%, aber die Existenz einer Kundenkarte besagt noch nicht, dass man das Potenzial an Daten auch nutzt.

Die erhobenen Informationen müssen zudem keineswegs allein aus internen Quellen stammen. Der beste Beleg sind Käuferreichweiten und Bedarfsdeckungsquoten, die ausschließlich über Haushaltspanels ermittelt werden können. Insoweit weisen die Antworten auch hinzugekaufte Informationen aus.

Auch sei die Anmerkung erlaubt: Über eine Information zu verfügen heißt noch lange nicht, dass sie auch sinnvoll verwendet wird. Der Wunsch nach mehr und besserer Information muss zum einen zur Einführung entsprechender Systeme und Kennzahlen führen, zum anderen stellt dies noch nicht ihre Nutzung im Tagesgeschäft sicher. Hier ist das Controlling gefordert, nicht nur Informationssysteme aufzubauen, sondern auch die Informationsversorgung zu garantieren (vgl. Schröder, Feller, Rödl 2003).



**Abbildung 6: Die Versorgung mit Transaktionsdaten im deutschen Handel**

## 3.2 Die Nutzung anonymer Transaktionsdaten

Transaktionsdaten aus einem Scannerkassensystem zeichnen sich durch die umfassende Abdeckung einzelner Einkaufsstätten aus. Daher sind sie vor allem geeignet, auf der Basis einzelner Bons oder über eine Gruppe von Bons die Umsatzstruktur eines Handelsbetriebes zu beschreiben. Je nach Fragestellung bietet es sich an, in einer *kundenorientierten Analyse* Einzelbons oder Bonsummenklassen gegenüberzustellen oder in einer *sortimentsorientierten Analyse* einzelne Produkte oder Warengruppen (Categories) zu betrachten. Wenn derartige Auswertungen z.B. ergeben, dass ein Händler seine Umsätze eher mit kleinen als mit großen Bons tätigt, ist weiter zu untersuchen, ob diese Umsätze aus Laufkundschaft oder Ergänzungskäufen resultieren. Stellt sich heraus, dass viele Kunden diesen Händler lediglich zur Deckung ihres Ergänzungsbedarfs nutzen, ihm aber in kaum einer Warengruppe Kompetenz zuschreiben (Analyse auf Warengruppenebene), sollte das Angebotskonzept überdacht werden.

Umsatzanalysen auf Artikelebene sind eine Grundlage für Listungsentscheidungen, indem die einzelnen Artikel nach der Höhe ihres Umsatzanteiles aufgereiht und alle Artikel unterhalb eines bestimmten Umsatzanteiles ausgelistet werden. Weitere Fragestellungen sind *zeitbezogene Umsatzanalysen*, um auf Schwankungen von Wochen-, Monats- oder Jahresumsätzen reagieren zu können. Sind zusätzlich Informationen zur Spanne verfügbar, können die Bruttoerträge betrachtet werden.

### 3.2.1 Messung von Verbundeffekten

Verwendungs- bzw. anlassorientierte Analysen folgen dem Motto: „Zeige mir, was mich interessiert!“. Dagegen arbeiten potenzialorientierte Methoden wie das Data Mining nach dem Motto: „Finde heraus, was interessant ist!“. Zum Data Mining gehören Methoden, die in großen Datenmengen Hypothesen über nicht-triviale Zusammenhänge aufdecken. Data Mining ist ein Sammelbegriff für verschiedene Verfahren, zu denen Assoziationsregeln, Entscheidungsbaumverfahren, Genetische Algorithmen, Clusteranalyse, Faktoranalyse, Regressionsanalyse, Diskriminanzanalyse, K-Nächste-Nachbarn, und Neuronale Netze gezählt werden (vgl. Wiedmann, Buckler ,Buxel 2001, 27 f.).

Eine erste Methode für die Analyse anonymer Transaktionsdaten ist die *Assoziationsanalyse*. Sie deckt in Datenbeständen das gemeinsame Auftreten von Objekten auf und liefert damit Regeln. Es handelt sich um ein Verfahren der Verbundanalyse, das mit drei Kennzahlen arbeitet (vgl. Decker, Schimmelpfennig 2002): Relevanz (Support), Konfi-

denz (Confidence) und Abweichung (Lift). Die Relevanz ist der Anteil der Transaktionen, die beide Itemmengen A und B enthalten, an allen Transaktionen der Menge W:

$$\text{sup}(A \Rightarrow B) = \frac{|\{T \in W \mid (A \cup B) \subseteq T\}|}{|W|}$$

Die Konfidenz einer Regel  $A \Rightarrow B$  ist der Anteil der Transaktionen, die beide Itemmengen A und B enthalten, an all jenen Transaktionen der Menge W, die zumindest den Regelrumpf A beinhalten:

$$\text{conf}(A \Rightarrow B) = \frac{|\{T \in W \mid (A \cup B) \subseteq T\}|}{|\{T \in W \mid A \subseteq T\}|}$$

Der Lift einer Regel gibt an, um welchen Faktor der Regelkopf B unter der Voraussetzung des Eintretens von A häufiger auftritt als in der gesamten Datenbasis.

Die Konfidenz ist ein Maß, das die Existenz asymmetrischer Verbundbeziehungen aufdeckt. Dies ist für die Sortimentsgestaltung nicht unwichtig, da man erfährt, von welchem Produkt eine Verbundwirkung ausgeht. Die Ergebnisse liefern Antworten auf zahlreiche Fragen, wie etwa:

- Welche Artikel einer Warengruppe werden zusammen gekauft?
- Von welchem Artikel geht die Verbundbeziehung aus?
- Welche Artikel werden zusammen mit Sonderangeboten gekauft?
- Welche Artikel einer Warengruppe werden mit Artikeln anderer Warengruppen gekauft?
- Welche Artikel finden sich auf Bons mit hohen und niedrigen Summen?

Assoziationsanalysen können nicht nur Beziehungen zwischen Artikeln aufdecken, sondern zwischen sämtlichen von dem Anwender definierten Objekten. Die Erkenntnisse der Assoziationsanalysen unterstützen Entscheidungen über den Einsatz absatzpolitischer Maßnahmen. Soweit dem Händler bekannt ist, zwischen welchen Produkten Verbundbeziehungen bestehen, kann er gezielt die verbundkaufauslösenden Artikel fördern. Dies ist besonders hilfreich bei der Gestaltung von Rabattaktionen. Derzeit vermitteln viele Aktionen, insbesondere im Bereich von Coupons, den Eindruck, dass das Gießkannenprinzip vorherrscht und nicht das Wissen um das Kaufverhalten der Kunden. Gleichzeitig lassen sich Streuverluste bei solchen Artikeln vermeiden, zwischen denen keine oder nur geringe Verbundwirkungen bestehen. Wurde bisher vermutet, dass be-

stimmte Maßnahmen in einer Warengruppe zu hohen Verbundeffekten führen und entkräften die Assoziationsanalysen diese Vermutung, ergeben sich – wie die Praxis zeigt – erhebliche Einsparpotenziale.

### 3.2.2 Charakterisierung von Käufen

Sofern man wissen möchte, wie sich bestimmte Gruppen von Objekten, z.B. Bons oder Kunden, zueinander verhalten, hilft die *Clusteranalyse*. Clusteranalysen können nach beliebig wählbaren Kriterien Segmente in den Datenbeständen bilden. Zu unterscheiden sind clusterbildende und clusterbeschreibende Variablen. Die clusterbildenden Variablen sind diejenigen Kriterien, nach denen homogene Gruppen gebildet werden. Mit den beschreibenden Variablen lassen sich diese Gruppen charakterisieren, was darüber hinaus natürlich auch mit den clusterbildenden Kriterien möglich ist. Clusteranalysen auf der Basis anonymer Daten können u.a. folgende Fragen beantworten:

- Wie verteilen sich die Artikel auf die einzelnen Warenkörbe?
- Welche Warengruppen werden zusammen gekauft?
- Welche Produkte / Warengruppen werden an welchen Tagen gekauft?
- Wie unterscheiden sich kleine von großen und sehr großen Bons?
- Wie lassen sich die Käufe einer bestimmten Warengruppe charakterisieren?

Durch Antworten auf diese Fragen lernt ein Händler die Struktur seiner Käufe wesentlich besser kennen. Das Beispiel in *Abbildung 7* verdeutlicht, welche Erkenntnisse sich aus der Clusteranalyse mit anonymen Transaktionsdaten ziehen lassen. Die clusterbildende Variable ist der Wertanteil, gemessen in der Zahlungswährung, den die Warengruppe „Wein“ am gesamten Einkauf hat. Als clusterbeschreibende Variablen wurden die in eckigen Klammern stehenden Variablen verwendet. Die Ergebnisse zeigen, dass Kunden, die einen hohen Wertanteil in der Warengruppe „Wein“ auf ihren Bons verzeichnen,

- öfter freitags und samstags in diesem Geschäft einkaufen als die übrigen Kunden,
- überdurchschnittlich hohe Anteile an Frische und Grundnahrungsmitteln aufweisen,
- eine überdurchschnittlich hohe Bonsumme haben,
- überdurchschnittlich viele Artikel kaufen etc.

Teilweise handelt es sich um nahe liegende Ergebnisse. Dies ist aber oft erst dann erkennbar, wenn bestimmte Teilergebnisse bekannt sind. So kann es wenig überraschen, dass die Bonsumme überdurchschnittlich hoch ist, wenn man auf die überdurchschnitt-

lich hohe Zahl an Artikeln auf den Bons verweist. Ausgehend von diesem Ergebnis sind weiter gehende Analysen sinnvoll, etwa auf Warengruppen- und Artikelebene. Diese Informationen sind hilfreich, um Entscheidungen über verschiedene Instrumente des Handelsmarketing zu treffen, angefangen bei der Sortimentszusammensetzung über die Platzierung von Artikeln bis hin zur Steigerung von Verbundkäufen bei denjenigen Weinkäufern, die bestimmte Warengruppen bislang vernachlässigen. Unter anderem lassen sich Couponaktionen gezielt auf bestimmte Kauftypen (nicht Käufer) richten, von denen bekannt ist, wann sie welche Artikel kaufen.

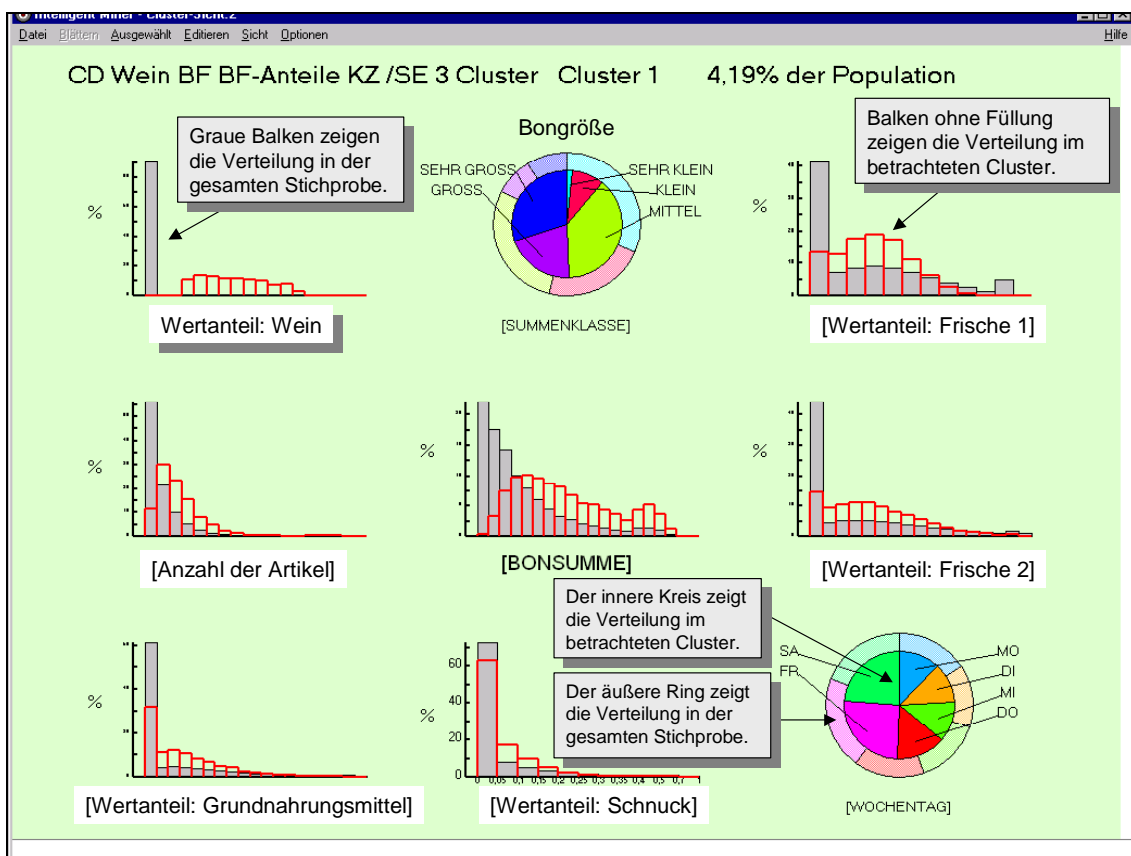


Abbildung 7: Ergebnis einer Clusteranalyse mit anonymen Transaktionsdaten

### 3.3 Die Nutzung pseudonymer Transaktionsdaten

Assoziationsanalysen mit anonymen Daten müssen sich auf Querschnittbetrachtungen (zeitpunktbezogene Analysen) beschränken. Dagegen erlauben pseudonyme Daten *Sequenzanalysen*, d.h. Längsschnittbetrachtungen (zeitraumbezogene Analysen). Pseudonyme Transaktionsdaten gestatten es, alle Käufe eines Kunden oder eines Haushaltes im Zeitablauf zu verknüpfen. Sequenzielle Muster sind also nichts anderes als über mehrere



Zeitpunkte reichende Assoziationen. Sequenzanalysen können z.B. folgende Fragen beantworten:

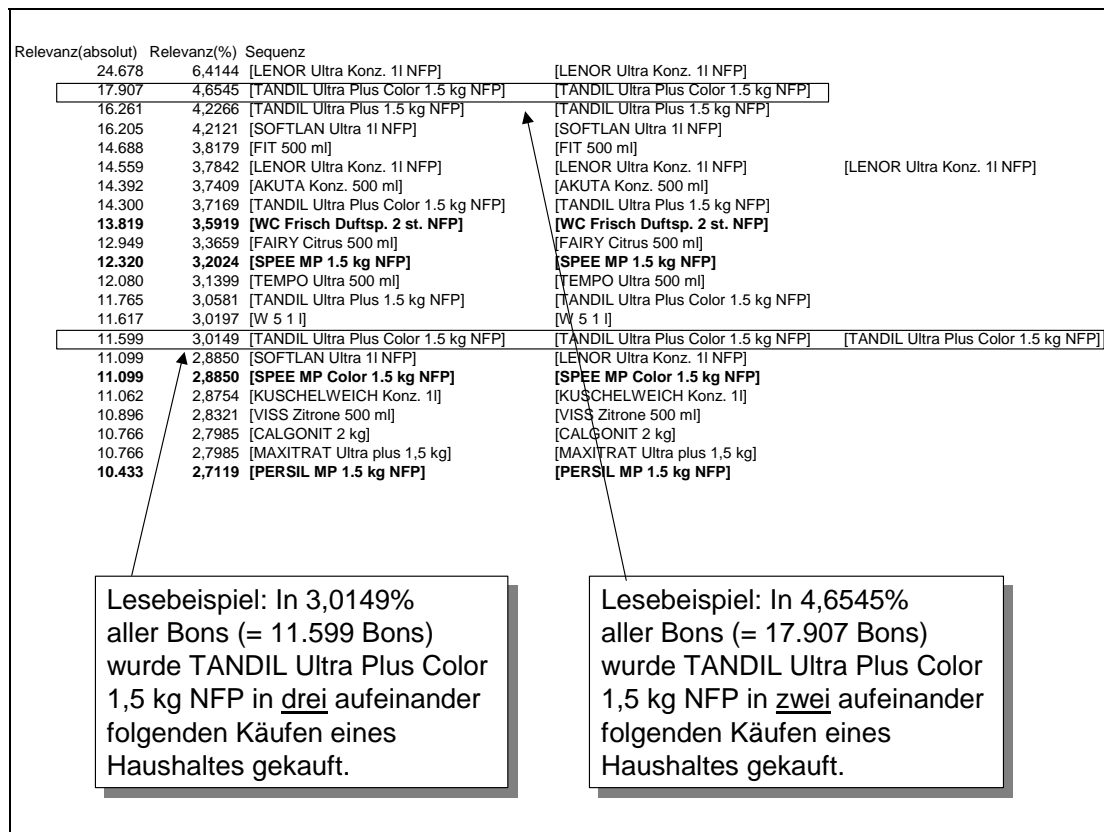
- Welche Artikel werden wie oft in aufeinander folgenden Einkäufen gekauft?
- Wie hoch ist die Markentreue bestimmter Artikel?
- Wie reagieren die Kunden im Zeitablauf auf Aktionen der eigenen Handelsunternehmung und der Konkurrenz?
- Wie ändert sich das Kaufverhalten im Zeitablauf?

Selbstverständlich lassen sich auf pseudonyme Daten auch alle jene Methoden anwenden, die für anonyme Daten möglich sind.

*Abbildung 8* zeigt Ausschnitte einer Sequenzanalyse. Die Ergebnisse helfen, Artikel mit geringer Markentreue aufzudecken, Produkte für Bonuspunkte- und Coupon-Programme auszuwählen, das Direktmarketing (Rückgewinnung von markenuntreuen Kunden) zu unterstützen und Aktionserfolge zu bewerten (Zusatzkäufe oder Verschiebung von Käufen). Damit zeigen sie auf, in welchen Sortimentsbereichen der Händler über einen bestimmten Zeitraum hinweg Stärken und Schwächen hat. Auch hier steht die Effizienzsteigerung der Marketinginstrumente im Vordergrund: förderungswürdige Maßnahmen erkennen, Streuverluste vermeiden.

Allerdings sollten weitere Informationen bereitgestellt werden. Es reicht nicht aus, vernachlässigte Artikel zu identifizieren und hierfür unmittelbar z.B. eine Couponaktion zu starten. Vielmehr ist zu bedenken, dass diese Aktionen zu Mindereinnahmen bei den treuen Kunden führen. Daher sind Kenntnisse über die Elastizität der Nachfrage in Abhängigkeit von der Höhe des Preisnachlasses ebenso erforderlich wie über den Verzicht auf Rohertrag bei den bislang abgesetzten Artikeln.

Transaktionsdaten aus dem Haushaltspanel liefern auch *einkaufsstättenübergreifende Analysen*, wie z.B. Käuferwanderungen und Verbundanalysen in Konkurrenzbetrieben. Damit wird dieses Instrument zur Konkurrenzbeobachtung genutzt. Die Ergebnisse geben Hinweise über die Erfolgswirksamkeit der eigenen Marketinginstrumente im Vergleich zu denen der Konkurrenz. Die Aussagekraft findet jedoch ihre Grenzen in der Stichprobe. National und überregional ausgerichtete Analysen werden auf einer halbwegs akzeptablen Datenbasis möglich sein, regionale und lokale Analysen dagegen nicht. Hier erreicht die Drill down-Analyse auch bei Fast Moving Consumer Goods (FMCG) schnell das Fallzahlenproblem.



**Abbildung 8: Ergebnisse einer Sequenzanalyse**

Pseudonyme Daten sind die Basis, um den *Customer Lifetime Value (CLV)* zu berechnen. Hierbei handelt es sich um den Wert, den ein Kunde für die gesamte Dauer seiner Kundschaft bei einer Unternehmung für diese darstellt (vgl. hierzu ausführlich Rödl 2003, S. 199 f.). Der CLV lässt sich aus der Herstellerperspektive (Produkte, Marken) und der Handlungsperspektive (Categories, Einkaufsstätten) ermitteln. Er ergibt sich aus der Differenz zwischen den durch einen Kunden erbrachten *Einzahlungen* an die Unternehmung und deren *Ausgaben* für diesen Kunden. Der CLV zeigt auf, welche Kunden(gruppen) wertvoll oder weniger wertvoll sind. Ebenso erfasst er, wie Kunden auf die verschiedenen Instrumente des Handelsmarketing, etwa die Facetten der Rabattpolitik, reagieren.

Die Berechnung des CLV kann mit Hilfe der Kapitalwertmethode erfolgen. Sie hat die nachfrager- und anbieterrelevanten Merkmale des Einzelhandels einzubeziehen. So berücksichtigt Formel (1) u.a. unterschiedliche Einkaufstage und damit auch die zeitliche Verteilung der einzelnen Warenkörbe eines Kunden:

$$CLV_{a0} = \sum_{t=0}^T \frac{\sum_{s=1}^S (E_{ast} - A_{ast}) - A_{at}}{(1+i)^t} \quad (1)$$

$CLV_{a0}$  = Customer Lifetime Value des Kunden a zum Zeitpunkt 0

$t \in (0, \dots, T)$  = Periode;  $T$  = Anzahl der Perioden, in denen der Kunde der Unternehmung erhalten bleibt

$s \in (1, \dots, S)$  = Einkaufstag;  $S$  = letzter Einkaufstag der betrachteten Periode

$E_{ast}$  = Einzahlung des Kunden a an Einkaufstag s der Periode t

$A_{ast}$  = Auszahlung für Kunden a, verursacht durch dessen Einkauf an Tag s der Periode t

$A_{at}$  = allgemeine Auszahlung für Kunden a in Periode t

$i$  = Kalkulationszinsfuß bezogen auf die Dauer einer Periode t

Berücksichtigt man lediglich die Einstandspreise der gekauften Produkte als kaufbezogene Auszahlung für einen Kunden, so bildet Formel (2) den Warenrohertrag ab.

$$E_{ast} - A_{ast} = \sum_{j=1}^J (p_{asjt} \cdot x_{asjt}) - \sum_{j=1}^J (ep_{asjt} \cdot x_{asjt}) \quad (2)$$

$j \in (1, \dots, J)$  = Produkte im Sortiment des Händlers

$p_{asjt}$  = Verkaufspreis des Produktes j für den Kunden a zum Einkaufszeitpunkt s in Periode t

$x_{asjt}$  = Menge des Produktes j zum Zeitpunkt des Einkaufs s des Kunden a in Periode t

$ep_{asjt}$  = Einstandspreis für Produkt j zum Zeitpunkt des Einkaufs s des Kunden a in Periode t

Die kundenbezogenen Auszahlungen [ $A_{ast}$ ] eines Händlers können vielfältiger sein und z.B. Auszahlungen für Direct-Mails und individuelle Beratung berücksichtigen. Alle nicht-kundenspezifischen Ausgaben wie z.B. Ladenmiete werden vernachlässigt, da sie unabhängig von den Kunden anfallen. In dem Wert eines Warenkorbes lässt sich auch erfassen, welche Arten von Preisnachlässen der Kunde in Anspruch genommen hat. Der CLV erfasst damit die Reaktionen der Kunden auf individuell oder pauschal ausgerichtete Verkaufsförderungsmaßnahmen und bildet dadurch verschiedene Kundentypen ab. So lässt sich z.B. aufdecken, wie häufig und in welchem Umfang Kunden einen ausschließlich an sie ausgegebenen Coupon genutzt haben.

Der absolute CLV sagt noch nicht abschließend etwas über den Wert eines Kunden aus. Liegt z.B. der CLV des Kunden A in einem Warenhaus bei 86.000 € und der des Kunden B bei 59.000 €, so bedeutet dies nicht ohne weiteres, dass der Kunde A wertvoller ist als der Kunde B. Vielmehr sind weitere Merkmale zur Abgrenzung von Kundengruppen heranzuziehen, wie z.B. das Lebensalter, die Haushaltsgröße und die Kaufkraft. Erst diese Differenzierungen erlauben die Einteilung in wertvolle und weniger wertvolle

Kunden. Für das Handelsmarketing ergibt sich daraus die Forderung, Konzepte zu erarbeiten, um die Bindung der wertvollen Kunden zu sichern und um weniger wertvolle Kunden zu wertvolleren zu entwickeln oder aber zu „verabschieden“.

### **3.4 Die Nutzung personalisierter Transaktionsdaten**

Personalisierte Daten erlauben dieselben Auswertungsmethoden, wie sie bereits für pseudonyme und anonyme Daten vorgestellt wurden. Der Zusatznutzen für den Anwender besteht darin, dass er aufgrund der Kenntnis des Kunden diesen individuell mit Marketingmaßnahmen ansprechen kann, soweit dem keine rechtlichen Restriktionen entgegenstehen. Auf diese Weise lassen sich Produkte und Warengruppen fördern, treue Kunden belohnen und Maßnahmen zurücknehmen oder vermeiden, die ganz offensichtlich ineffizient sind.

Einige Beispiele: Kann festgestellt werden, dass ein Kunde entgegen der Erwartung (Haushalt mit Kleinstkindern) bestimmte Warengruppen (Babynahrung etc.) vernachlässigt, so bieten sich gezielte Informationen an, um auf die Leistungsfähigkeit in dieser Warengruppe aufmerksam zu machen. Auch lassen sich mit Fragen an die Adressaten mögliche Schwächen in dieser Warengruppe aufdecken. Dagegen sollte man Aktivitäten zurücknehmen, wenn etwa festgestellt wird, dass der Grad an Verbundkäufen in einer Warengruppe oder zwischen Warengruppen bereits ausgeschöpft ist, oder wenn nach einer Couponaktion der gewünschte Effekt der Nachfragesteigerung erreicht wurde. Personalisierte Transaktionsdaten sind ebenfalls geeignet, um die Einkaufsstättentreue durch zeitbezogene Rabatte (z.B. auf Jahresebene) zu steigern oder zu stabilisieren.

## **LITERATUR**

Decker, R.; Schimmelpfennig, H.: Alternative Ansätze zur datengestützten Verbundmessung im Electronic Retailing, in: Ahlert, D.; Olbrich, R.; Schröder, H. (Hrsg.): Jahrbuch zum Handelsmanagement 2002 - Electronic Retailing, Frankfurt am Main 2002, S. 193-212.

Ebbers, D.: Welche Daten können was im Category Management-Prozess?, in: Schröder, H. (Hrsg.): Category Management: Aus der Praxis für die Praxis - Konzepte - Kooperationen - Erfahrungen, Frankfurt am Main 2003, S. 141-156.

Knobloch, B.: Der Data-Mining-Ansatz zur Analyse betriebswirtschaftlicher Daten, Bamberger Beiträge zur Wirtschaftsinformatik, Nr. 58, Bamberg 2000.

- Meffert, H.: Marketingforschung und Käuferverhalten, 2. Aufl., Wiesbaden 1992.
- Rödl, A.: Der Customer Lifetime Value von Kunden des Lebensmittel-Einzelhandels als Steuerungsgröße im Category Management, in: Schröder, H. (Hrsg.): Category Management: Aus der Praxis für die Praxis - Konzepte - Kooperationen - Erfahrungen, Frankfurt am Main 2003, S. 197-217.
- Schäfer, A.: Resources of Information - Information Resources GfK GmbH, unveröffentlichte Schulungsunterlagen, o.O. 2002.
- Schröder, H., Feller, M., Rödl, A.: Leistungen des Controlling für eine kundenorientierte Sortimentsgestaltung im Lebensmittel-Einzelhandel, in: Graßhoff, J., Krey, A. (Hrsg.): Handelscontrolling - Neue Ansätze aus Theorie und Praxis zur Steuerung von Handelsunternehmen, 2. Aufl., Hamburg 2003 (im Druck).
- Schröder, H.; Feller, M.; Schuck, A.: Wie erfolgreich ist CRM tatsächlich? - Ergebnisse einer empirischen Untersuchung, in: Ahlert, D., Becker, J., Knackstedt, R., Wunderlich, M. (Hrsg.): Customer Relationship Management im Handel, Berlin, Heidelberg, New York 2002, S. 397-422.
- Wiedmann, K.-P.; Buckler, F.; Buxel, H.: Data Mining - ein einführender Überblick, in: Wiedmann, K.-P.; Buckler, F. (Hrsg.): Neuronale Netze im Marketing-Management, Wiesbaden 2001, S. 15-33.