

DeltaMaster clicks!

04/2015

Liebe Datenanalysten,

der große Vorrat an Analyseverfahren ist eine besondere Stärke von *DeltaMaster*. Dazu zählen, ganz offensichtlich, die Methoden im Modus *Miner* mit ihren speziellen Algorithmen, Heuristiken und Berichtsformaten. Daneben gibt es zahlreiche weitere Funktionen, die innerhalb von tabellarischen Berichten bereitstehen und neue Erkenntnisse liefern. In diese Kategorie fällt die Clusteranalyse. Sie fahndet nach ähnlichen Strukturen in den Berichtsdaten und hebt hervor, was auffällig ist. Dabei ist sie so einfach anzuwenden und geht so unauffällig zu Werke, dass wir uns fast gefragt haben, ob wir diese Ausgabe der *DeltaMaster clicks!* überhaupt damit füllen können – aber genau diese Unauffälligkeit soll auch einmal auffallen dürfen.

Herzliche Grüße

Ihr Team von Bissantz & Company

So unterstützen wir Sie bei der Arbeit mit DeltaMaster

Online-Hilfe

Direkt in *DeltaMaster*:
Menü *Hilfe* oder Taste *F1*

Support-Hotline

support@bissantz.de
Tel. +49 911 935536-700

Newsletter

DeltaMaster clicks!
Jeden Monat. Archiv und Abo:
www.bissantz.de/clicks

DeltaMaster deltas!
Zu jedem neuen Release.
Abo auf Anfrage.

Blogs

Me, myself and BI – Bissantz denkt nach
blog.bissantz.de

Bella berät – die meisten Diagramme sind für die Katz
www.bella-beraet.de

Die Newsletter und die Artikel dieser beiden Blogs sind in die Online-Hilfe integriert.

Auf die Würfel, fertig, los – wie wir Ihren Daten Beine machen
crew.bissantz.de

Bissantz forscht – Neues aus unseren Laboren
forschung.bissantz.de

Schulungen

Gut 100 Schulungstage rund um *DeltaMaster* und Microsoft SQL Server/Analysis Services.
www.bissantz.de/Schulungen

Veranstaltungen

Erleben Sie *DeltaMaster* live – zum Beispiel auf Kundentreffen, Fachseminaren, Informationstagen, Kongressen oder Messen.
www.bissantz.de/Veranstaltungen



DeltaMaster-Forum am 24. März 2015 in München

Vor der spektakulären Kulisse der Allianz-Arena ging es um die Frage, was zu tun ist, damit Berichte Wirkung entfalten. Bei diwa Personalservice tun sie das, wie Matthias Budde und Uwe Prell (links im Bild) in ihrem Vortrag eindrücklich belegen konnten. Vielen Dank Ihnen beiden! In der Bildmitte: Dr. Nicolas Bissantz, rechts: Prof. Dr. Dr. Gerhard Roth. Das nächste *DeltaMaster*-Forum findet am 10. Juni in Frankfurt statt, dann mit dem *DeltaMaster*-Kunden Mars Drinks.
www.bissantz.de/DeltaMaster-Forum

Kniff des Monats Ähnliche Wertverteilungen mit der Clusteranalyse aufdecken

Zur genaueren Untersuchung von Pivot- und Kreuztabellen bietet *DeltaMaster* eine Reihe von Schnellanalysen an. Diese wirken wie ein guter Assistent: Sie nehmen die Tabelle unter die Lupe und heben automatisch hervor, was bemerkenswert ist. Eine dieser Schnellanalysen ist die Clusteranalyse. Sie identifiziert ähnliche Objekte und fasst sie zu Gruppen zusammen. Das benötigt man beispielsweise bei der Kundenprofilierung oder Marktsegmentierung – oder einfach, um die Objekte in der Tabelle einmal auf Ähnlichkeiten abzuklopfen.

Ein Cluster (auf Deutsch etwa: Gruppe, Anhäufung, Ansammlung, Ballung) soll dabei so definiert sein, dass die Objekte in diesem Cluster möglichst ähnlich zueinander sind – und möglichst verschieden von den Objekten in anderen Clustern. Es sollen also die Cluster in sich möglichst homogen sein, damit die enthaltenen Objekte gut zu charakterisieren sind; untereinander sollen sie möglichst heterogen sein, um trennscharf Handlungen folgen zu lassen.

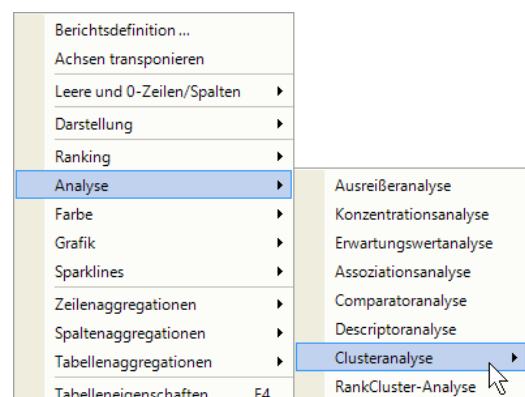
Absatz	...	Kunde	...	Süd	Nord	Ost	West
Produkt	...	+	+	+	+	+	+
Arcade AE 66	26,1 %	●	67,2 %	●	3,1 %	·	3,7 %
Precisio JK	30,2 %	●	67,3 %	●	1,5 %	·	1,0 %
EF Drehst. Presa	28,9 %	●	59,9 %	●	7,0 %	*	4,2 %
EF Konferenz	29,4 %	●	64,7 %	●	3,8 %	·	2,0 %
EF Besucherst. MO	24,4 %	●	57,5 %	●	13,7 %	*	4,4 %
Nova B	27,0 %	●	68,2 %	●	3,0 %	·	1,8 %
Arcade AE 44	15,7 %	●	63,3 %	●	12,2 %	*	8,7 %
Precisio LF	9,5 %	●	63,3 %	●	17,3 %	*	9,9 %
Hansen ZZ	10,1 %	*	56,9 %	●	23,7 %	●	9,3 %
EF Drehst. Ergo	1,6 %	·	89,5 %	●	3,4 %	·	5,5 %
EF Besucherst. MP	4,2 %	·	92,7 %	●	0,8 %	·	2,2 %
Fronius Basic			100,0 %	●			

Die Clusteranalyse in Pivot- und Kreuztabellen funktioniert in *DeltaMaster* besonders einfach und benötigt keine Parameter oder Vorwissen – die Ergebnisse sprechen für sich.

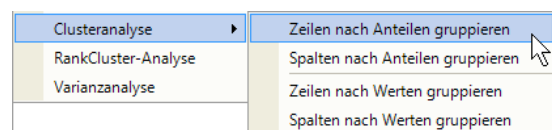
Analyse starten

Angeboten wird die Clusteranalyse in zwei Modulen: in Pivottabellen und in der *Kreuztabellenanalyse*. Für Pivottabellen ist der Modus *Pivotizer*, *Analyzer* oder *Miner* erforderlich, für die *Kreuztabellenanalyse* der Modus *Analyzer* oder *Miner*. Von der Methodik her funktioniert die Analyse in beiden Modulen gleich, lediglich die Analyseketten-technik ist der *Kreuztabellenanalyse* vorbehalten.

Zusammen mit den anderen Schnellanalysen ist die *Clusteranalyse* im Kontextmenü der Tabelle zu finden, und zwar im Abschnitt *Analyse*.



Im Untermenü wählen Sie aus, welche Objekte *DeltaMaster* gruppieren soll (Zeilen oder Spalten) und nach welchen Kennzahlen (Anteilen an der Zeilen- bzw. Spaltensumme oder absoluten Werten).



Das war's schon! Sofort startet *DeltaMaster* die Berechnung und versucht, Zeilen oder Spalten zusammenzufassen, bei denen sich die Werte oder Anteile ähnlich auf die Elemente der jeweils anderen Achse verteilen.

Ballungsräume

Als Beispiel haben wir in unserer Referenzanwendung „Chair“ die Produkt- und die Kundendimension gegenübergestellt, eine durchaus typische Ausgangslage für Clusteranalysen. Zur Verdeutlichung der Wertverhältnisse sind global skalierte Kreise eingeblendet (Kontextmenü, Abschnitt *Grafik*; die Skalierung legt man in den *Tabelleneigenschaften* auf der Registerkarte *Grafik* (1) fest). Diese Darstellung mag zunächst etwas überraschen: In den meisten Berichten verwenden wir Balken oder Säulen, weil man Längen und Höhen besser abschätzen kann als Flächen. Kreise haben jedoch den Vorteil, dass sie sich gut in zwei Richtungen vergleichen lassen, vertikal und horizontal, zeilen- und spaltenweise, und genau das ist hier gefragt. Produkte und Regionen sind „gleichberechtigte“ Merkmale, Kreise somit geeignet zu signalisieren, welche Kombinationen von Bedeutung sind („Kreise können kreuzweise“, siehe blog.bissantz.de/kreuzkreise). Aber trotz der passenden Grafiken: Ähnliche Nachfrageprofile sind so nicht auf einen Blick zu erkennen.

Absatz Produkt	Kunde			
	Süd	Nord	Ost	West
Arcade AE 44	27.505	111.091	21.479	15.325
Arcade AE 55	12.177	25.253	12.088	4.280
Arcade AE 66	8.099	20.852	957	1.140
Arcade AE 77	6.087	2.456		196
Discus Comp	155		41	
Discus Office	1.755			
EF Besucherst. MO	3.884	9.153	2.181	699
EF Besucherst. MP	11.930	260.745	2.212	6.258
EF Drehst. Ergo	608	33.628	1.279	2.055
EF Drehst. Nova	133.410	142.125	1.433	1.322
EF Drehst. Presa	37.090	76.865	8.958	5.425
EF Konferenz	10.069	22.144	1.308	680
Ergoplus Basic		10		
Ergoplus Nova	3.358	44.154	10.570	151.333
Ergoplus Senso	6.422	10.004		
Hansen 10	63.583	103.353	42.556	88.440
Hansen ZZ	5.664	31.860	13.271	5.218
Nova B	27.497	69.440	3.087	1.868
Nova C	8.200	15.333	6.211	3.250
Precisio JK	171.741	383.358	8.673	5.813
Precisio LF	11.162	74.469	20.292	11.699
Precisio LM	8.126	7.407	2.261	

Die Clusteranalyse ändert das sofort. In der nebenstehenden Abbildung haben wir *DeltaMaster* die *Zeilen nach Anteilen* gruppieren lassen. Die Tabellenansicht wird dabei automatisch umgeschaltet, um die Anteile an der Zeilensumme anzuzeigen (anstelle der absoluten Ergebnisse). Die Clusteranalyse sortiert die Zeilen so um, dass die ähnlichen Elemente zusammenstehen; eine eventuelle Sortierung der Achse wird dabei ignoriert. Mit der schwarzen Umrandung markiert *DeltaMaster* die Cluster. Im Beispiel wurden deren fünf gefunden – und der optische Eindruck bestätigt den statistischen Befund.

Absatz Produkt	Kunde			
	Süd	Nord	Ost	West
Arcade AE 66	26,1 %	67,2 %	3,1 %	3,7 %
Precisio JK	30,2 %	67,3 %	1,5 %	1,0 %
EF Drehst. Presa	28,9 %	59,9 %	7,0 %	4,2 %
EF Konferenz	29,4 %	64,7 %	3,8 %	2,0 %
EF Besucherst. MO	24,4 %	57,5 %	13,7 %	4,4 %
Nova B	27,0 %	68,2 %	3,0 %	1,8 %
Arcade AE 44	15,7 %	63,3 %	12,2 %	8,7 %
Precisio LF	9,5 %	63,3 %	17,3 %	9,9 %
Hansen ZZ	10,1 %	56,9 %	23,7 %	9,3 %
EF Drehst. Ergo	1,6 %	89,5 %	3,4 %	5,5 %
EF Besucherst. MP	4,2 %	92,7 %	0,8 %	2,2 %
Ergoplus Basic		100,0 %		
EF Drehst. Nova	47,9 %	51,1 %	0,5 %	0,5 %
Ergoplus Senso	39,1 %	60,9 %		
Arcade AE 55	22,6 %	46,9 %	22,5 %	8,0 %
Nova C	24,9 %	46,5 %	18,8 %	9,9 %
Discus Comp	78,9 %		21,1 %	
Discus Office	100,0 %			
Arcade AE 77	69,7 %	28,1 %		2,2 %
Precisio LM	45,7 %	41,6 %	12,7 %	
Ergoplus Nova	1,6 %	21,1 %	5,0 %	72,3 %
Hansen 10	21,3 %	34,7 %	14,3 %	29,7 %

- Im ersten Cluster liegen Produkte mit hohem Umsatz im Norden, mittlerem Umsatz im Süden und niedrigem Umsatz im Osten und Westen – jeweils gemessen am Gesamtumsatz mit diesem Produkt, also der Zeilensumme.
- Im zweiten Cluster liegen Produkte mit hohem Umsatz im Norden und etwa gleichen, eher niedrigen Umsätzen in den übrigen Regionen.
- Im dritten Cluster liegen Produkte, die praktisch nur im Norden „laufen“.

- Im vierten Cluster liegen Produkte, deren Umsatz überwiegend aus dem Norden, zu einem guten Teil aber auch aus dem Süden stammt; die anderen Regionen sind zu vernachlässigen.
- Im fünften Cluster liegen Produkte, bei denen auch der Osten und Westen einen beträchtlichen Umsatzanteil haben.

Außerhalb der Cluster, am Ende der Tabelle, stehen einige Einträge, die zu keinem der anderen Muster und auch nicht zueinander so recht passen wollen. Deshalb sind sie keinem Cluster zugeordnet.

Wenn man mit absoluten Werten anstatt mit Anteilen rechnet, ergibt sich ein anderes Bild. Beide Analysevarianten haben ihre Berechtigung. Häufig konzentriert man sich auf die absoluten Werte und dabei wiederum auf die Cluster mit hohen Werten. Aber auch relative Betrachtungen sind sinnvoll und interessant. Beispielsweise mag man für Produkte mit ähnlicher Absatzstruktur gleiche Richtlinien in der Versandlogistik entwickeln, auch wenn die Umsätze in unterschiedlichen Größenordnungen liegen.

Absatz Produkt	Kunde			
	Süd	Nord	Ost	West
Arcade AE 77	6.087	2.456		196
Precisio LM	8.126	7.407	2.261	
Discus Comp	155		41	
Discus Office	1.755			
EF Besucherst. MO	3.884	9.153	2.181	699
ErgoplusBasic		10		
ErgoplusSenso	6.422	10.004		
Arcade AE 55	12.177	25.253	12.088	4.280
Arcade AE 66	8.099	20.852	957	1.140
EF Drehst. Ergo	608	33.628	1.279	2.055
EF Konferenz	10.069	22.144	1.308	680
Hansen ZZ	5.664	31.860	13.271	5.218
Nova C	8.200	15.333	6.211	3.250
EF Drehst. Presa	37.090	76.865	8.958	5.425
Nova B	27.497	69.440	3.087	1.868
Precisio JK	171.741	383.358	8.673	5.813
Precisio LF	11.162	74.469	20.292	11.699
EF Drehst. Nova	133.410	142.125	1.433	1.322
Arcade AE 44	27.505	111.091	21.479	15.325
EF Besucherst. MP	11.930	260.745	2.212	6.258
ErgoplusNova	3.358	44.154	10.570	151.333
Hansen 10	63.583	103.353	42.556	88.440

Wird die Pivottabelle oder die Kreuztabelle mit berechneter Clusteranalyse gespeichert, so ist sie auch im Modus *Viewer* verfügbar. Beim Aktualisieren des Berichts oder etwa bei einem Wechsel der *Sicht*, zum Beispiel zu einem anderen Monat, berechnet *DeltaMaster* die Clusteranalyse automatisch neu.

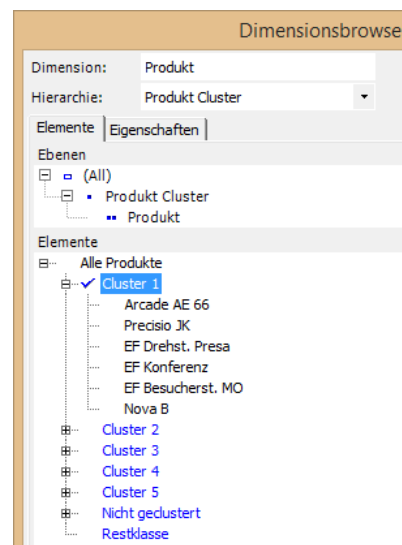
Unter der Haube

Das Verfahren der Clusteranalyse ist in *DeltaMaster* so angelegt, dass es für den Anwender möglichst einfach zu bedienen und robust gegenüber unterschiedlichen Datenkonstellationen ist. Wer es genau wissen will: Die Ähnlichkeit von Objekten wird über den sogenannten Euklidischen Abstand ermittelt, die Gruppierung erfolgt nach dem Average-Linkage-Verfahren. Zusätzlich sorgt ein Regelwerk dafür, dass die Analyse auch unter widrigen Umständen brauchbare Ergebnisse liefert. Beispielsweise werden Sie es dank adaptiver Schwellwerte nur selten erleben, dass kein Cluster oder nur ein Cluster gefunden wird, der alle Objekte enthält.

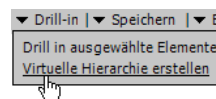
Analysekettentechnik: gefundene Cluster weiterverwenden

Eine schnelle Clusteranalyse kann für viele Zwecke schon genügen – als Hinweis auf Zusammenhänge in den Daten, die man ohne Systemunterstützung kaum hätte finden können. Manchmal will man aber mehr: nämlich mit den ermittelten Gruppen weiter analysieren und die Cluster in anderen Berichten und Analysen aufgreifen.

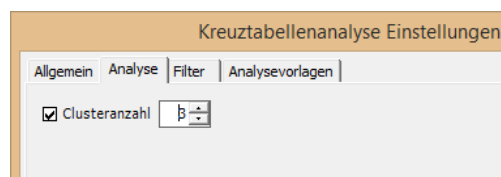
Im Modus *Miner* ist dies in der *Kreuztabellenanalyse* möglich: Die gefundenen Gruppierungen lassen sich als sogenannte virtuelle Hierarchie speichern. Diese wiederum steht im *Dimensionsbrowser*, in *Pivottabellen*, *Analyseverfahren* usw. genauso zur Verfügung wie die Dimensionen, die *DeltaMaster* aus der Datenbank übernommen hat. Veränderungen an der Datenbank sind jedoch nicht erforderlich, die Hierarchie wird allein in *DeltaMaster* verwaltet. Die virtuellen Elemente heißen zunächst „Cluster 1“, „Cluster 2“ usw. Im *Dimensionsbrowser* lassen sie sich umbenennen (Kontextmenü oder Taste *F2*). Elemente, die in keinen Cluster fallen, werden dem virtuellen Element „Nicht geclustert“ zugeordnet. Bei Bedarf legt *DeltaMaster* außerdem eine „Restklasse“ an, als Sammelposten für Elemente, die in der aktuellen Tabellenansicht ausgeschlossen sind, zum Beispiel leere oder nach Analysewerten ausgefilterte Zeilen oder Spalten. Einzelheiten finden Sie in den *DeltaMaster deltas!* 5.5.5, Punkt 10.



Um eine *virtuelle Hierarchie* zu erstellen, rufen Sie den entsprechenden Befehl im Menü *Drill-in* (Fenster *Analyse*) auf. Voraussetzung ist, dass die Clusteranalyse berechnet ist und Cluster gefunden wurden.



Vor allem für die Weiterverwendung als virtuelle Hierarchie kann es sinnvoll sein, eine feste *Clusteranzahl* vorzugeben. Diese lässt sich in den *Einstellungen* der *Kreuztabellenanalyse* eintragen (siehe *DeltaMaster deltas!* 5.5.6, Punkt 7).



Universell einsetzbar

Die Clusteranalyse können Sie auf beliebige Pivot- und Kreuztabellen ansetzen. Prädestiniert ist sie für Merkmale wie Kunden, Regionen, Produkte, Materialien und Ähnliches – Merkmale, deren Kombination gut zu interpretieren ist und eine plausible Gruppendefinition erlaubt. Auch verschachtelte Achsen sind möglich, etwa, wenn man Artikel und Ausstattungsvarianten gemeinsam betrachtet. Nur in einigen Anwendungsgebieten wird man von Clusteranalysen Abstand nehmen, aus inhaltlichen Gründen. So ist es wenig sinnvoll, Konten, GuV-Zeilen oder etwa die Analysewerte eines Kennzahlenschemas zu gruppieren. Auch entlang der Zeitdimension wird man im Allgemeinen nicht nach Clustern suchen.