

Nachfolgend zunächst einige generelle Anmerkungen zur Frage der Datenaufbereitung, die vielleicht als Checkliste sinnvoll ist:

- 1.) Erhalten und prüfen der Daten auf ihre Plausibilität
- 2.) Prüfen der Daten auf Inkonsistenz

Multi-Source-Probleme

Ursache: überlappende, widersprüchliche bzw. inkonsistente Daten, die unabhängig voneinander in den entsprechenden Quellen erzeugt und gespeichert wurden

a) Hauptproblem: überlappende Daten

- Gängige Bezeichnungen: *Duplikate*, *Merge/Purge-Problem*, *Object Identity Problem*
- Beschreibung einer Instanz der realen Welt durch mehrere Datensätze unterschiedlicher Quellen
- Oft nur teilweise Redundanz (einzelne Attribute, nur in Teilmenge der Datenquellen)

b) Unterschiedliche Repräsentationen der Instanzdaten

- verschiedene Wertebereiche (z.B. *Geschlecht* = {1,2} vs. *Gender* = {m,w})

verschiedene Einheiten (z.B. *Verkauf* in DM vs. *Verkauf* in TDM)

Euro / DM !!

verschiedene Genauigkeiten

c) inkonsistentes Timing: unterschiedliche Änderungsstände der Quelldaten

generelle Unterschiede zwischen Bewegungsdaten und Stammdaten

Frage: Will ich die Kundenumsätze anhand der Zugehörigkeit zu einer bestimmten Kundengruppe analysieren, zu dem der Umsatz entstanden ist, also die Rechnung geschrieben wurde oder möchte ich die Daten zur Kundengruppe sehen, so wie die Zuordnung per dato besteht? Gehen die Quelldaten auf Bewegungsdaten, auf Stammdaten oder evtl. sogar auf eine Mischung beider zurück?

d) unterschiedliche Aggregationsstufen der Quelldaten

3.) Host-Daten enthalten Zahlenwerte

Diese sind jedoch nicht immer sinnvolle Begriffe, wenn es um die Aufbereitung um Daten geht. Z. B. sind dort die Vertreter-Nummern hinterlegt, jedoch nicht die Namen der Vertreter. Die Daten enthalten einen Code für eine Kundengruppe, nicht jedoch den Namen der Kundengruppe u.s.w.

Vor der weiteren Bearbeitung sollte immer zunächst eine Prüfung der Vollständigkeit sinnvoller bzw. verständlicher Begriffe stehen.

Beispiel:

Es ist ärgerlich, nach getaner Arbeit festzustellen, dass das Resultat verständlicher wäre, wenn nicht die Datei/Liste lediglich eine Vertreter-Nummer statt den Namen der betreffenden Kollegin oder des Kollegen enthält.

Je nach Umfang der Quelldaten sollten diese Ergänzungen über Excel-Funktionen wie SVERWEIS() u.ä. bzw. über eine Access-Abfrage realisierbar sein.

4.) Auch die fertigen Daten möglichst immer auf Plausibilität prüfen

Grundsätzlich sollte keine Auswertung herausgegeben werden, die man nicht vorher auf Plausibilität geprüft hat. Gerade in Zeiten der Umstellung der Landeswährung auf Euro werden einige Zahlenbändiger ein flaes Gefühl in der Magengegend bei der Abgabe mancher Berichte haben.

Erstellung einer Pivot-Tabelle

Gehen wir davon aus, dass die Quelldaten lt. Checkliste (siehe oben) geprüft sind und bestanden haben. Die Quelldaten liegen in Form einer Excel-Tabelle vor. Achten Sie darauf, dass die Feldnamen (Spaltenköpfe) mit eindeutigen Namen betitelt sind und dass diese Namen in Ihrer Tabelle nicht doppelt vorkommen.

Beispiel:

	A	B	C	D	E	F	G	H	I
1	KDNR	KG	KG-Name	MOOD	PLZ_ORT	STRASSE	Gebiet	Vertreter	Umsatz
2	1	1	Karstadt	KARSTADT HBGHARB	21073 HAMBURG 90 HARBURG	SCHLOSSMUEHLENDAMM 2	200	Müller	130.000
3	2	1	Karstadt	KARSTADT HBMOEN	20095 HAMBURG 1	MOENCKEBERGSTR. 16	200	Müller	130.229
4	3	1	Karstadt	KARSTADT BREMEN	28195 BREMEN 1	OBERNSTR. 5-33	200	Müller	130.541
5	4	3	Domaene	DOMAENE HAMBURG	37178 HARDEGSEN	SCHULSTR. 2	200	Müller	130.870
6	5	3	Domaene	DOMAENE HAMBURG WAIMANN	37178 HARDEGSEN	SCHULSTR. 2	200	Müller	131.633

Verwenden Sie Feldnamen, in denen keine Sonderzeichen oder Leerzeichen vorkommen (z.B. «PLZ_ORT» nicht „PLZ / ORT“). In den neuen Excel Versionen geht die Verwendung solcher Schreibweisen zwar meistens ohne Probleme vonstatten, trotzdem sollte auch deshalb darauf verzichtet werden, weil die Daten vielleicht später in ein anderes Format konvertiert werden sollen, das vielleicht weniger tolerant sind.

Beginnen wir mit der Aufbereitung einer Pivot-Tabelle. Dazu rufen wir das Menü »Daten« auf und dort den Punkt «PivotTable und PivotChart-Bericht».

Beispiel:



Nun wird der Pivot-Tabellen-Assistent erscheinen und wir folgen den Vorschlägen des Assistenten, d.h. wir lassen die Voreinstellungen wie sie sind. Die Vorgaben sind beide zutreffend. Wir wollen eine vorhandene Excel-Liste- bzw. Datenbank analysieren und im ersten Schritt eine Pivot-Tabelle erstellen.

Beispiel:



Nun wird der komplette Datenbereich des aktiven Tabellenblattes markiert und die Bezüge in das Formular übernommen. Auch diese Einträge können Sie einfach übernehmen. Sollten Sie einmal innerhalb einer Excel-Tabelle verschiedene Datenbanken analysieren wollen, so kann es hilfreich sein, den verschiedenen Bereichen Listen/Datenbanken) sinnvolle Namen zuzuweisen, damit kein „Durcheinander“ entsteht.

Beispiel:

WPK	MOB	PLZ	ORT	STRASSE
bst	KARSTADT-HOHKAMP	21072	HAMBURG 93 HAMBURG	SCHLOSSMÜHLENDAMM 2
bst	KARSTADT-HOBOMICH	21095	HAMBURG 1	MOENCKEBOERSTR. 18
bst	KARSTADT-BRGMEN	21195	BREMEN 1	OBERNGSTR. 5-23
wee	DOMAENE HAMBURG	21176	HARDELOSEN	SCHULSTR. 2
wee	DOMAENE HAMBURG HW	21176	HARDELOSEN	SCHULSTR. 2
in	FAMILIA WILDESHAU	21793	WILDESHAUEN	AHLHORNERSTR. 67
is	FAMILIA WILDESHAU	21777	WILDESHAUEN	GRLEPPENBUCHENWEGE
es	GLOBUS STRALSUND	18442	GROSS LUEDERSHAGENWENDORF	ALBERT-SCHNEITZER-STR.
es	GLOBUS HROSTORF			
es	GLOBUS KALTENB.			
	BAHR LUEDERSTR. 10			11
	BAHR HAMBURG			23
	BAHR LUEDERSTR. 11			ALL STR. 85
	BAHR ROSTOCK			ZEN 1
	BAHR FLANDERS			FER STR. 11
	BAHR SCHWERN			
	BAHR HOBOMICH			
	BAHR ELMHORN	25336	ELMHORN	WICHTERSTR. 16
	BAHR RASDORF	24223	KIEL-RASDORF	MERTENTHALER STR. 34
	BAHR HOBOMICH	21073	HAMBURG 93 HAMBURG	BUNTEHUEBEN STR. 62
	BAHR BREMEN	28307	BREMEN	STREESCHMIDTSTR. 68
	BAHR BREMEN-HWEN	27676	BREMEN-HWEN	STREESCHMIDTSTR. 120

Hierzu gehen Sie so vor, dass Sie die Datenbank zunächst komplett markieren (also nur den Bereich, in dem sich Daten befinden und zwar incl. der Feldnamen). Jetzt wählen Sie «Einfügen» und dann «Name» «Definieren». Nun können Sie in den unterlegten Bereich einen „sinnvollen“ Namen (z.B. „Vertreterumsatz“) eingeben. Andere Datenbank-Bereiche Ihrer Tabelle können Sie in gleicher Weise mit einem Namen versehen und sorgen hiermit für mehr Übersicht.

Weiter geht es mit unserer Pivot-Tabelle:

Der Schritt 3 des Assistenten schlägt vor, dass die spätere Pivot-Tabelle in einem neuen Tabellenblatt dargestellt wird. Dies ist sicher immer dann sinnvoll, wenn die Pivot-Tabelle einiges Volumen haben wird bzw. viele Daten in einer weniger hohen Dichte verarbeitet werden. Sie können also auch die Pivot-Tabelle im gleichen Blatt erscheinen lassen.

Beispiel:



Jetzt gibt es zwei Möglichkeiten, mit der Erstellung der eigentlichen Pivot-Tabelle fortzufahren: Aus der Excel 8-Zeit wäre jetzt eigentlich direkt der Punkt «Layout» fällig. Unter Excel 9 ist das noch einfacher. An dieser Stelle können wir nun mit «Fertig stellen» die vorbereitenden Arbeiten abschliessen.

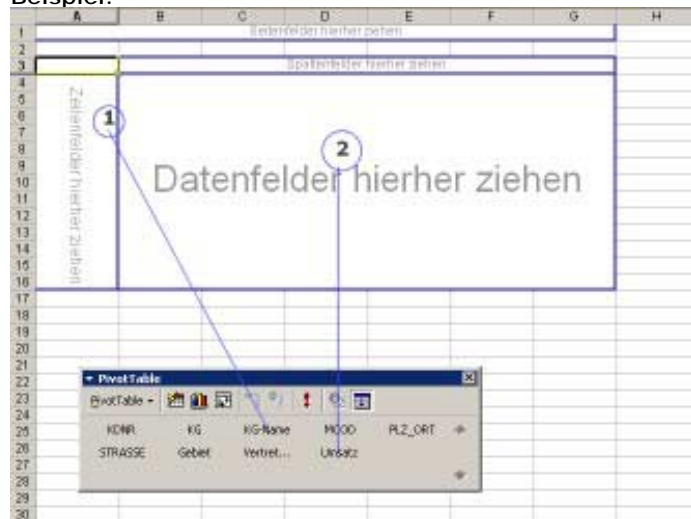
Vorgeblendet wird nun eine Layout-Ansicht der eigentlichen Pivot-Tabelle. Sie können anhand der im unteren Teil der ebenfalls erscheinenden separaten Symbolleiste die Felder mit gedrückter Maustaste per Drag and Drop in die bezeichneten Bereiche ziehen (Seitenfelder, Spaltenfelder, Zeilenfelder und Datenfelder). Die Bedeutung dieser Begriffe erschließt sich in dem Moment, in dem Sie testweise einmal zwei Felder herüberziehen. In unserem Beispiel wäre dies KG-Name (Kundengruppe-Name) und Umsatz. Ziehen Sie das Feld «KG-Name» in den Bereich „Zeilenfelder“ und das Feld «Umsatz» in den Bereich „Datenfelder“. Praktischer Weise beginnen Sie immer zunächst entweder mit den Spalten- oder mit den Zeilenfeldern. Erst danach sollten Sie ein Datenfeld (Felder mit zu berechnenden Werten) in den Datenbereich ziehen.

Arbeiten und Behandlung von Pivot-Tabellen

von Horst Schulte – Aug. 2001

4/7

Beispiel:



Ergebnis:

Summe - Umsatz	
KG-Name	Ergebnis
Bahr	5020261
Domaene	262503
Familia	265639
Globus	402843
Hombach	152754
Karstadt	390770
Obi	462282
Raab Kärcher	153498
Gesamtergebnis	7110550

Anhand unseres Beispielen sehen Sie nun die Gesamtumsätze nach «KG-Name» (Kundengruppe).

An diesem simplen Beispiel erkennen Sie sofort, auf welche schnelle und komfortable Weise Sie mit diesem mächtigen Analyse-Tool Daten auswerten und aufbereiten können.

Jetzt wäre es doch interessant, wenn wir innerhalb einer Kundengruppe auch noch sehen könnten, wie hoch die Umsätze unserer Vertreter mit diesen Kundengruppen wären. Also ziehen wir mit gedrückter linker Maustaste auch das Feld «Vertreter-Name» in unser Pivot-Tabellen-Layout und zwar zwischen die beiden bereits vorhandenen Felder «KG-Name» und «Summe-Umsatz», welches als Ergebnis angezeigt wird.

Ergebnis:

Summe - Umsatz	Vertreter	Ergebnis
Bahr	Kleier	2051382
	Müller	1095893
	Münch	604291
	Schmidt	1268595
Bahr Ergebnis		5020261
Domaene	Müller	262503
		262503
Familia	Müller	265639
		265639
Globus	Müller	402843
		402843
Hombach	Münch	152754
		152754
Karstadt	Müller	390770
		390770
Obi	Münch	462282
		462282
Raab Kärcher	Münch	153498
		153498
Gesamtergebnis		7110550

Diese Pivot-Tabelle ist noch nicht formatiert und sieht noch wenig „ungeschliffen“ aus. Doch dazu später mehr.

Ziehen Sie jetzt bitte mit der Maus das Feld «Umsatz» aus der immer noch sichtbaren Pivot-Table-Symboleiste nochmals in den Datenbereich, also neben das bereits vorhandene Feld «Umsatz». Was machen wir jetzt mit zwei Umsatzsummen-Feldern? Zum ersten sehen wir, dass neben dem Feld «Umsatz Summe» nun das Feld «Umsatz Summe2» erscheint. Das ändern wir zunächst einmal, damit wir nachher mit den Bezeichnungen nicht durcheinander kommen.

Hierzu überschreiben wir einfach das Feld «Umsatz Summe2» mit «%-Anteil». Da haben wir auch gleich den Grund für die Einführung dieses zweiten Umsatzsummen-Feldes. Stellen Sie den Mauszeiger auf das gerade geänderte Feld «%-Anteil». Wir manipulieren dieses nun, in dem wir in der Pivot-Table-Symboleiste das Symbol «Feldeinstellungen» drücken (zweites von rechts oben). Jetzt erscheinen alle möglichen Optionen, die wir für dieses Feld benutzen können. Die Voreinstellung für Ziffern (um die es sich beim Feld Umsatz ja auch handelt, ist »Summe«).

Folgende Zusammenfassungsoptionen sind möglich:

Funktion	Ergebnis
Summe	Die Summe der Werte. Standardfunktion für numerische Quelldaten.
Anzahl2	Die Anzahl von Elementen. Die Zusammenfassungsfunktion Anzahl arbeitet in gleicher Weise wie die Tabellenfunktion ANZAHL2. Anzahl2 ist die Standardfunktion für nichtnumerische Daten.
Mittelwert	Der Mittelwert der Werte.
Maximum	Der höchste Wert.
Minimum	Der niedrigste Wert.
Produkt	Das Produkt der Werte.
Anzahl	Die Anzahl von Zeilen, die numerische Daten enthalten. Die Zusammenfassungsfunktion Anzahl arbeitet in gleicher Weise wie die Tabellenfunktion ANZAHL.
Standardabweichung (Stichprobe)	Eine Schätzung der Standardabweichung einer Population, wobei alle zusammenzufassenden Daten als Stichprobe dienen.
Standardabweichung (Grundgesamtheit)	Die Standardabweichung einer Population, wobei alle zusammenzufassenden Daten (Grundgesamtheit) die Population darstellen.
Varianz (Stichprobe)	Eine Schätzung der Varianz einer Population, wobei alle zusammenzufassenden Daten als Stichprobe dienen.
Varianz (Grundgesamtheit)	Die Varianz einer Population, wobei alle zusammenzufassenden Daten (Grundgesamtheit) die Population darstellen.

Mit diesem (oberen) Teil wollen wir uns jetzt aber nicht beschäftigen, sondern nur mit dem unteren „Daten anzeigen als...“.

Dort finden wir den Begriff „%- der Spalte“. Genau diesen wählen wir aus und bestätigen die Wahl mit der Maustaste. Sofort sehen wir, wie Excel die entsprechende Berechnung vornimmt und die %-Anteil der Spalte zum Gesamtwert errechnet.

Arbeiten und Behandlung von Pivot-Tabellen

von Horst Schulte – Aug. 2001

6/7

Beispiel:

	A	B	C	D
1	Seitenfelder hierher ziehen			
2				
3			Daten	
4	KG-Name	Vertreter	Summe - Umsatz	%-Anteil
5	Bahr	Kleier	2051382	28,85%
6		Müller	1095893	15,41%
7		Münch	604391	8,50%
8		Schmidt	1268595	17,84%
9	Bahr Ergebnis		5020261	70,60%
10	Domaene	Müller	262503	3,69%
11	Domaene Ergebnis		262503	3,69%
12	Famila	Müller	265639	3,74%
13	Famila Ergebnis		265639	3,74%
14	Globus	Müller	402843	5,67%
15	Globus Ergebnis		402843	5,67%
16	Hornbach	Münch	152754	2,15%
17	Hornbach Ergebnis		152754	2,15%
18	Karstadt	Müller	390770	5,50%
19	Karstadt Ergebnis		390770	5,50%
20	Obi	Münch	462282	6,50%
21	Obi Ergebnis		462282	6,50%
22	Raab Karcher	Münch	153498	2,16%
23	Raab Karcher Ergebnis		153498	2,16%
24	Gesamtergebnis		7110550	100,00%

Nun sieht unsere Pivottable bezüglich ihrer Aussage schon ganz gut aus. Was jedoch, wenn wir alle der gleichen Meinung wären und nun unser Chef meint, er möchte gerne die Darstellung so haben, dass der Vertreter und nicht die Kundengruppe «KG-Name» vorne steht. Mal sehen, wie lange wir benötigen, um die Auswertung entsprechend zu schwenken (drehen).

Das war wohl keine Sekunde!

Ergebnis:

		Daten	
Vertreter	KG-Name	Summe - Umsatz	%-Anteil
Kleier	Bahr	2051382	28,85%
Kleier Ergebnis		2051382	28,85%
Müller	Bahr	1095893	15,41%
	Domaene	262503	3,69%
	Famila	265639	3,74%
	Globus	402843	5,67%
	Karstadt	390770	5,50%
Müller Ergebnis		2417648	34,00%
Münch	Bahr	604391	8,50%
	Hornbach	152754	2,15%
	Obi	462282	6,50%
	Raab Karcher	153498	2,16%
Münch Ergebnis		1372925	19,31%
Schmidt	Bahr	1268595	17,84%
Schmidt Ergebnis		1268595	17,84%
Gesamtergebnis		7110550	100,00%

Dazu habe ich im Layout unserer Pivot-Tabelle mit der Maus das Feld «Vertreter» angeklickt und bei gedrückter Taste nach links (also vor das Feld «KG-Name») geschoben. Das war's.

Zwei kleine Tipps am Rande:

1.) Klicken Sie mal auf eine Zahl im Datenbereich («Summe Umsatz» oder «%-Anteil»). Dann werden alle zu diesem Datensatz gehörigen Werte in einer neuen Tabelle aufgelistet. Schöne Sache. Denken Sie aber bitte auch daran, dass, falls Sie eine solche Auswertung mal an einen Kunden schicken sollten, er die Basisdaten auf diese Weise sehen kann. Selbst wenn Sie eine Pivottable aus Ihrem Arbeitsblatt in eine neue Tabelle kopieren, so sind diese Daten trotzdem abrufbar.

2.) Sie können selbstverständlich auf einem Blatt verschiedene Pivottabellen darstellen. In unserem letzten Beispiel hätte ich die Ursprungstabelle (also die der Chef so nicht wollte), einfach kopieren können und daneben wieder einfügen können. Die beschriebene Änderung hätte ich dann sofort in der Kopie vornehmen können. So hätte ich beide Darstellungen in einem Blatt. Nur so als Idee.

Ich habe eine solche Lösung mit mehreren gleichartigen Pivottabellen nebeneinander realisiert, um Ranglisten abzubilden.

Beispiel:

Rang	Glasart	Wert
1	TRKL	100
2	AQU	95
..

Rang	Farbe	Wert
1	weiss	500
1	silber	250
..

Rang	Größe	Wert
1	90-90	500
1	80-80	450
..

Formatierungsbeispiele können Sie aus der Datei entnehmen, die Sie hier downloaden können.

Über ein Feedback würde ich mich sehr freuen.

hs@Horst-schulte.de