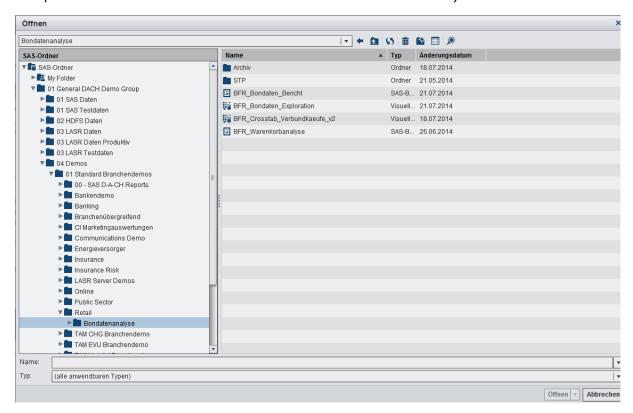
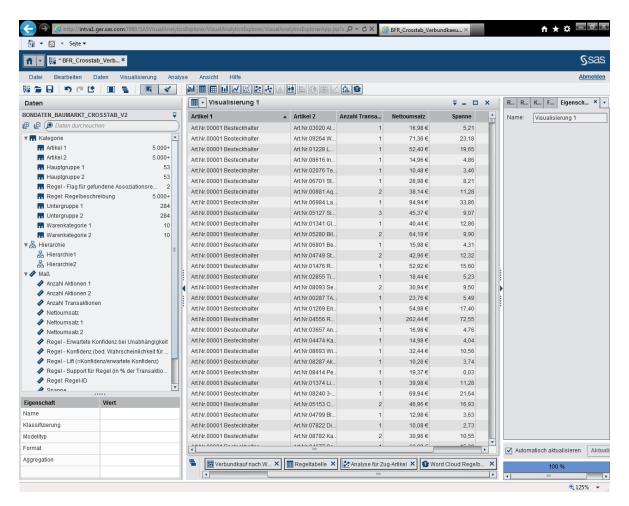
Demoskript Bondaten-Analyse Story (Teil 2):

Um zu erkennen, welche Artikel häufiger gemeinsam gekauft werden – und damit erste Anhaltspunkte für Promotion-Bündel zu erhalten, greifen wir auf eine weitere Exploration namens BFR_CROSSTAB_VERBUNDKAEUFE_V2 zurück. Du findest sie im Ordner 01 General DACH Demo Group – 04 Demos – 01 Standard Branchendemos – Retail - Bondatenanalyse:



Öffne diese Exploration. Sie enthält dieselben Bondaten wie bisher gesehen – nur etwas anders aufbereitet (Datenquelle: Bondaten_Baumarkt_Crosstab_v2). Es wurde jede paarweise Kombination von Artikeln gebildet und dann ausgezählt, in wie vielen Transaktionen (Warenkörbe) diese Kombination vorkommt und welcher Nettoumsatz und welche Spanne (Rohertrag) damit insgesamt erzielt wurden. Eine solche Aufbereitung ist mit SQL-Mitteln (COUNT Funktion!) grundsätzlich einfach möglich. Die zugrunde liegende Tabelle kann ebenso wie Ursprungstabelle persistent in den Hauptspeicher geladen werden und steht dann für die Fachanwender jederzeit für die interaktive Analyse zur Verfügung (Vorteil der In-Memory-Technologie). Du kannst exemplarisch den Aufbau zeigen, in dem Du eine Tabelle als Visualisierungsobjekt erzeugst und die beiden Artikel 1, Artikel 2 sowie Anzahl Transaktionen, Nettoumsatz und Spanne als Rolle zuweist:

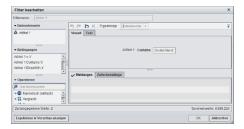


Die beiden Spalten Artikel 1 und Artikel 2 definieren jeweils eine Kombination von Artikeln. Zusätzlich zu den Artikelnamen (einschließlich Nummer) existieren auch noch Spalten für die Zugehörigkeit in der Sortimentshierarchie, also Warenkategorie 1, Hauptgruppe 1, Untergruppe 1 für Artikel 1, Warenkategorie 2, Hauptgruppe 2, Untergruppe 2 für Artikel 2. Mit diesen Spalten kann man bei Bedarf über die Sortimentshierarchie filtern. Es wurden auch entsprechende Hierarchien bereits angelegt. Auch die Maße wie Anzahl Aktionen, Nettoumsatz und Spanne wurden entsprechend berechnet. Schließlich existieren noch Spalten, die aus einem Data Mining Ansatz (Assoziationsregeln) kommen, doch dazu später mehr.

Zeige zunächst, wie ein Category Manager für einen ganz bestimmten Artikel die jeweiligen Verbundartikel findet, also diejenigen, die besonders häufig mit diesem Artikel gemeinsam gekauft werden. In unserem Fall weiß der Category Manager, dass zum Referenzzeitraum Mai vor der letzten Fussball-EM ein Artikel "Deutschlandfahne" als Promotion-Artikel vorhanden war. Er weiß aber nicht mehr genau, wie die Artikelnummer war. Für diesen Artikel möchte er nun die Verbundkauf-Kandidaten finden, die damals besonders häufig zusammen mit der Deutschlandfahne verkauft wurden.

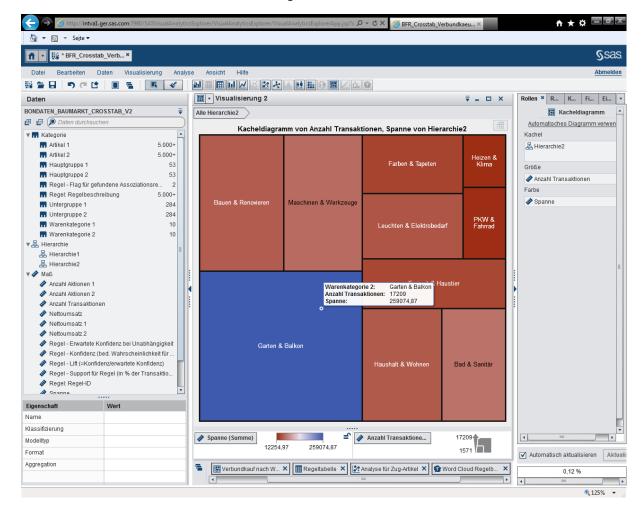
Am besten eignet sich dafür ein Kacheldiagramm-Objekt.

- 1. Erzeuge ein leeres Kacheldiagramm-Objekt
- 2. Wähle den Reiter Filter und ziehe die Spalte Artikel 1 in den Abschnitt Lokaler Filter.
- 3. Über die rechte Maustaste wähle das Kontextmenü Filter bearbeiten
- 4. Wähle im sich öffnenden Dialogfenster unter Bedingungen die Option Artikel 1 contains ,X'.
- 5. Tippe die Zeichenkette "Deutschland"

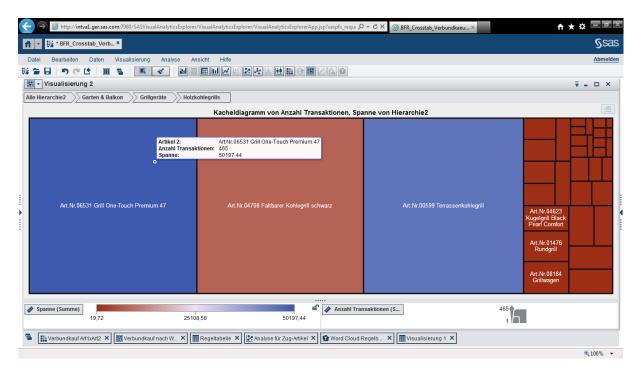


Klicke auf OK. Die Daten werden nun gefiltert, d.h. nur Zeilen mit Deutschlandfahne als Artikel 1 werden für die Analyse berücksichtigt.

Ziehe nun die Spalten Hierarchie 2 (nicht Hierarchie 1!), Anzahl Transaktionen und Spanne in die Rollen für Kachel, Größe und Farbe. Dein Diagramm müsste dann ähnlich wie dieses aussehen:



Im resultierenden Diagramm erkennt man sofort, dass die Kategorie Garten & Balkon heraussticht. Hier finden sich offenbar besonders geeignete Verbundartikel. Du kannst nun die Ansicht maximieren (Menü Ansicht oder entsprechendes Menüleistensymbol) und dann in die Kategorie hineindrillen, bis Du zur letzten Ebene (Artikel 2) kommst. Hier findest Du als Ergebnis, dass zwei bestimmte Holzkohle-Grills offenbar am meisten mit dem Artikel Deutschlandfahne verkauft wurden:



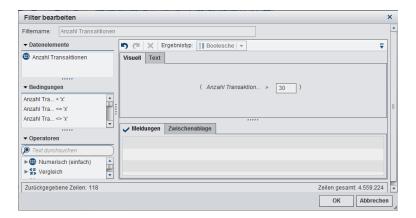
Übrigens: Das Diagramm existiert auch schon als vorgefertigte Visualisierung "Verbundkauf nach Warenkategorie".

Zwischenfazit 1:

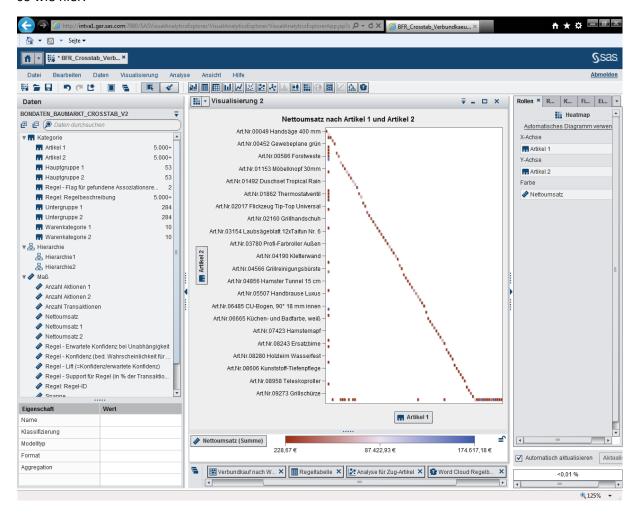
Für eine neue Aktion im Rahmen der WM 2014 – würde man, sofern man wieder eine Deutschlandfahne als Aktionsartikel heranzieht, schauen, ob man entsprechende Holzkohlegrill-Artikel im Sortiment hat und diese verwenden. Was aber, wenn der Category Manager gar nicht, welcher Artikel – so wie hier die Deutschlandfahne – damals im Angebot war? Dann kann er auch nicht danach filtern, sondern muss sich der Sache erst noch nähern?

Dafür eignet sich eine Heatmap-Visualisierung. Hier kann man grundsätzlich alle paarweisen Artikelkombinationen darstellen und über die Farbgebung die Anzahl der Transaktionen oder Umsatz bzw. Rohertrag darstellen. Allerdings sollte man sinnvollerweise filtern, weil sonst die Anzahl der darzustellenden paarweisen Artikelkombinationen die Möglichkeiten der Heatmap sprengt und das Auffinden relevanter Datenpunkte auch nicht mehr praktikabel wäre.

- Erzeuge ein leeres Heatmap-Visualisierungsobjekt.
- Ziehe die Spalte Anzahl Transaktionen in den Lokalen Filter, wähle *Filter bearbeiten* und setze die Filterbedingung "Anzahl Transaktionen > 30 " so wie hier

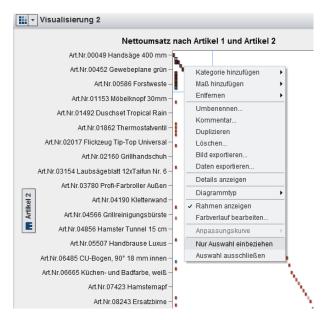


Ziehe dann die Spalten Artikel 1, Artikel 2 und Nettoumsatz in die Rollen X-Achse, Y-Achse und Farbe so wie hier:

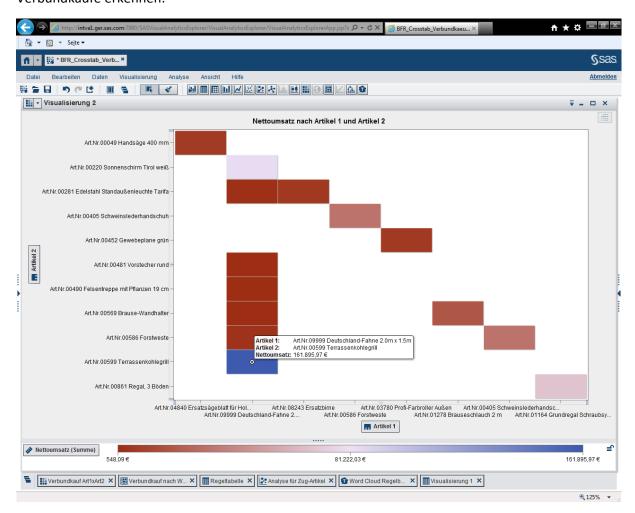


Du kannst anschließend die Ansicht wieder maximieren, um die Lesbarkeit zu erhöhen.

Prinzipiell kann man über die Färbung nun erkennen, für welche Kombinationen überhaupt Verbundkäufe existieren und wie hoch der Nettoumsatz ist (je blauer, desto höher). Allerdings sprengt macht die hohe Anzahl der Artikel die Lesbarkeit aber noch unübersichtlich. Hier hilft aber die In-Memory-Technologie, da ich über einen einfachen Weg aus der Grafik heraus die Daten filtern kann. Markiere einen beliebig kleinen Ausschnitt links oben in der Grafik und wähle über die rechte Maustaste das Kontextmenü *Nur Auswahl einbeziehen*.



Anschließend wird die Grafik neu aufgebaut. Nun kann ich die einzelnen Artikelkombinationen und Verbundkäufe erkennen:



Auch hier ist offenbar die Kombination Terassenkohlegrill x Deutschlandfahne besonders beachtenswert (in abgeschwächtem Maße auch der Sonnenschirm Tirol). (Tipp: Du kannst in den Eigenschaften der Heatmap die Option Achsenetiketten drehen verwenden, um die recht langen Artikelbeschreibungen schräg anzuordnen und damit etwas besser lesbar zu machen.)

In ähnlicher Weise könnte man jetzt den Filter wieder aufheben und in andere Bereiche der Heatmap hineinzoomen und diese "erforschen".

Zwischenfazit 2:

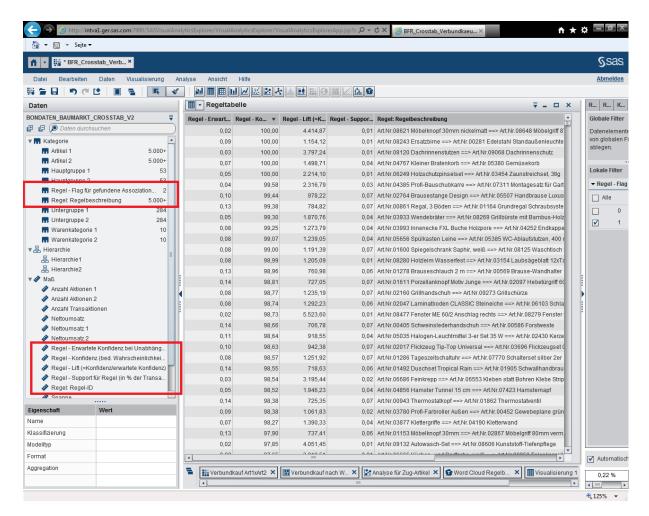
Über die Heatmap-Darstellung kann der Category Manager auch solche Verbundkaufeffekte erkennen, bei denen er nicht einen bestimmten Fokus-Artikel wie die Deutschlandfahne vorher filtert. Damit erhöht sich die Chance, auch überraschende Ergebnisse zu finden, die man je nach Kreativität neu auswerten kann.

Allerdings wird das iterative Hineinzoomen in die einzelnen Teile der Heatmap mit der Zeit etwas aufwändig. Besser wäre es, wenn der Category Manager automatisch auf bestimmte Effekte aufmerksam gemacht würde. Hier wird aber der Analyse-Umfang von Visual Analytics verlassen und stattdessen auf die Mustererkennung im Enterprise Miner zurückgegriffen. Der Enterprise Miner bringt Funktionalität mit, um sogenannte Assoziationsregeln der Form "Wenn A gekauft wurde, dann wurde auch B gekauft" automatisch aufzudecken. Dahinter stecken zwar prinzipiell einfache Häufigkeitsauszählungen, die man auch mit SQL-Skripten oder Abfragewerkzeugen machen könnte. Die schiere Anzahl der möglichen Kombinationen von Artikeln macht das aber in der Praxis unmöglich. Der Enterprise Miner bringt dafür einen Algorithmus (eine Variante des sog. A-Priori-Algorithmus) mit, der diese Aufgabe übernimmt. Es gibt nun zwei Möglichkeiten, diese Funktionalität einem Category Manager praxistauglich zur Verfügung zu stellen:

- (1) Automatisches Ausführen des Algorithmus im Batchlauf und Anreichern der Basis-Tabelle um die gefundenen Regeln und die zugehörigen statistischen Kennzahlen
- (2) Interaktives Ausführen des Algorithmus durch den Category Manager aus der gewohnten Oberfläche Visual Analytics heraus über das Vehikel eines SAS Stored Process (der die entsprechenden SAS Prozeduren zur Laufzeit ausführt und auch zur Laufzeit vom Anwender parametrisiert werden kann)

Zunächst wird die erste Variante vorgestellt.

In der Tabelle sind bereits Spalten enthalten, die im Batchlauf zur Tabelle hinzugefügt wurden und die im Rahmen einer Tabellenvisualisierung dargestellt werden können:



Es wurden hier insgesamt 10.000 Regeln automatisch gefunden (Spalte Regel ID enthält die fortlaufende Nummer) und an die Datentabelle angespielt. Die Interpretation ist wie folgt:

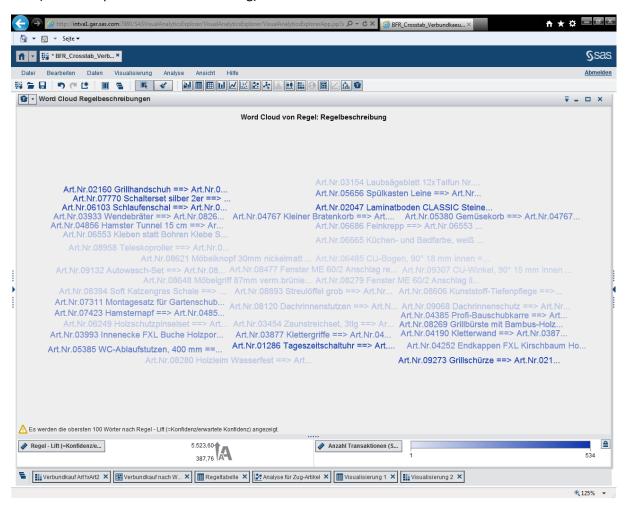
Flag für gefundene Assozationsregel	Nicht jede Artikelkombination enthält eine zugehörige Regel. Über diesen Flag kann man entsprechend die Tabelle filtern.
Regelbeschreibung	In der Form: Artikel A → Artikel B (d.h. wenn linke Regelseite erfüllt ist, dann folgt rechte Regelseite)
Konfidenz	Wahrscheinlichkeit dafür, dass wenn Warenkorb Artikel in der linken Regelseite enthält, auch Artikel der rechten Regelseite enthalten ist
Erwartete Konfidenz	Wahrscheinlichkeit für rechte Regelseite, wenn die beiden Kaufentscheidungen für die Artikel A und B nichts miteinander zu tun hätten, die Entscheidungen also unabhängig erfolgen würden
Lift	Quotient aus Konfidenz und erwarteter Konfidenz, also Faktor für das Maß der Stärke des Verbundkaufs
Support	Anteil der Warenkörbe, in denen diese Regel vorkommt.

Die Regeltabelle kann man nun durchforsten – und bei Bedarf auch noch mal filtern (z.B. Schwellwert für Support, Lift oder Konfidenz). Lesebeispiel:



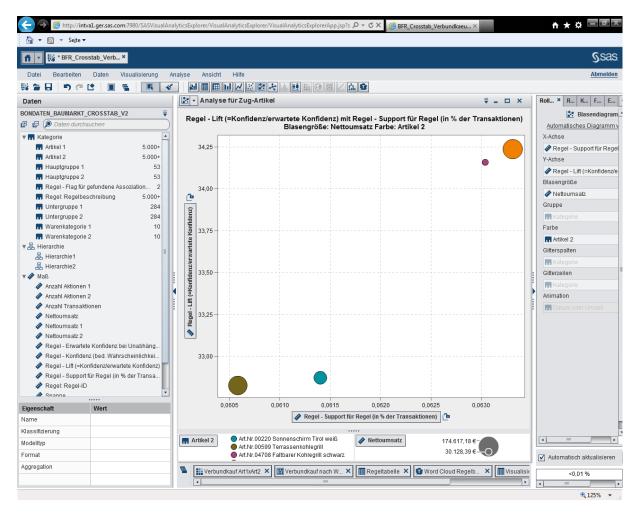
In 0,01% der Transaktionen (Spalte Support) wurde die Regel gefunden: Wenn der Artikel 06477 Fenster Anschlag Rechts gekauft wurde, dann wurde mit einer Wahrscheinlichkeit von 98,73 % auch der Artikel 06279 Fensteranschlag Links gekauft. Wenn die beiden Artikel-Kaufentscheidungen nichts miteinander zu tun hätten (was bei Fensteranschlägen unwahrscheinlich wäre), dann wäre die erwartete Wahrscheinlichkeit nur 0,02% gewesen. Deswegen ergibt sich ein extrem hoher Lift-Wert von 5.523 als Quotient. Die genannte Regel ist zwar statistisch gesehen sehr stark, aber kaum verwertbar, da sie keinen Überraschungseffekt wie "Bier & Pampers" birgt.

Man kann die Regeln auch anders darstellen, z.B. als Heatmap oder in Form einer Word-Cloud so wie hier (siehe entsprechende Visualisierung):



Eine alternative Darstellung ist ein Blasendiagramm, bei dem für einen bestimmten Artikel (Zug-Artikel) die Größen Lift und Support gegenübergestellt werden und die Blasengröße eine andere Dimension wie z.B. Nettoumsatz repräsentiert. Dabei muss man aber einerseits einen lokalen Filter auf den Zugartikel legen (hier Deutschlandfahne), andererseits einen Filter auf den Flag für Regeln. Andernfalls wird die Darstellung zu groß und unüberschaubar.

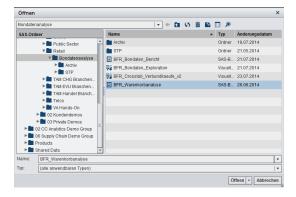
Auch hierfür gibt es bereits eine fertige Visualisierung:



Hier tauchen für den Artikel Deutschlandfahne ebenfalls die bekannten Zusammenhänge mit den Verbund-Artikeln Holzkohlegrills und Sonnenschirm Tirol auf.

Möchte man nicht auf vorberechnete Regeln zurückgreifen, sondern dem Category Manager als Fachanwender die Möglichkeit geben, die Regeln interaktiv zu finden und dabei auf bestimmte Parameter Einfluss zu nehmen, muss man einen entsprechenden Stored Process ausführen. Den kann man z.B. aus einem mit dem Designer erstellten Bericht heraus aufrufen.

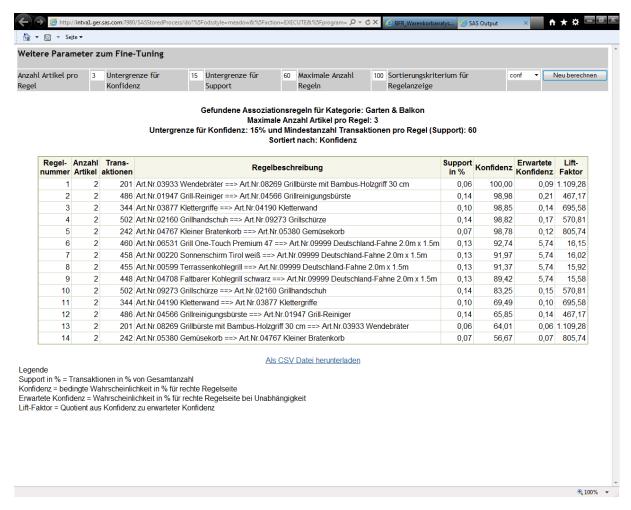
Um dies zu zeigen, verlasse die aktuelle Exploration und öffne über den Visual Analytics Hub den Bericht BFR Warenkorbanalyse im Retail-Ordner:



Der Bericht stellt zunächst die Umsatzverteilung nach Warenkategorien als Balkendiagramm dar. Hinter jedem Balken wird aber der Stored Process verlinkt und ausgeführt. Der Stored Process filtert dabei auf den jeweils markierten Balken.



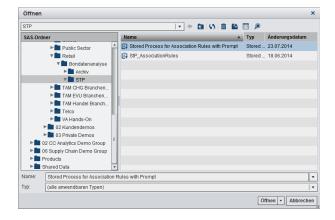
Klicke nun auf den Balken für Garten & Balkon. Der Stored Process wird nun ausgeführt. Die Rohdatentabelle wird gefiltert (nur Warenkategorie Garten & Balkon) und mit voreingestellten Parameterwerten ausgeführt. Das Ergebnis wird als HTML-Seite dargestellt.



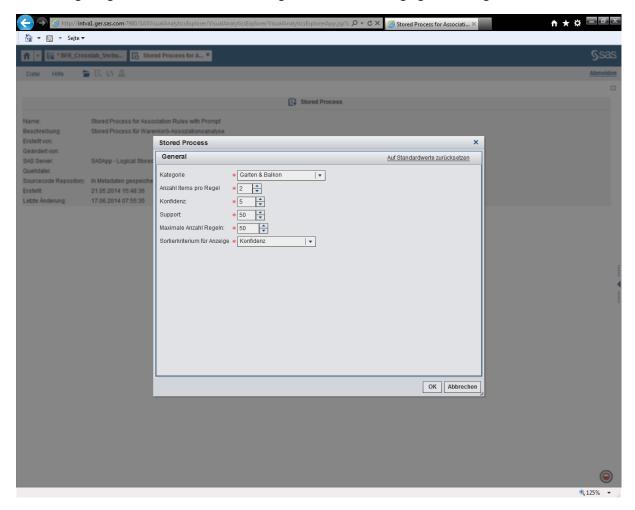
Über die Leiste am oberen Rand kann man die einzelnen Parameter noch mal ändern und über die Schaltfläche *Neu berechnen* den Stored Process neu ausführen. Außerdem kann man die gefundenen Ergebnisse als CSV-Datei herunterladen. Beim Klicken auf einen anderen Balken würden die gefundenen Regeln für die jeweilige Warenkategorie erzeugt.

Anhand der Tabelle kann ich nun wiederum durchforsten, welche der gefundenen Regeln interessant ist. Es gibt sicher auch visuell ansprechendere Darstellungen zur Aufbereitung der Ergebnisse, aber dieses Demo-Beispiel soll ja nur ein die grundsätzliche Machbarkeit der Integration analytischer Funktionalität andeuten.

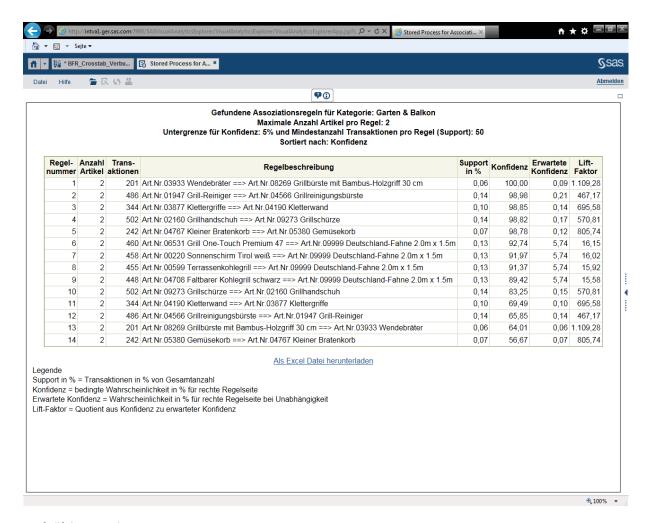
Ein alternativer Weg findet sich im Unterverzeichnis STP unter Bondatenanalyse.



Durch Klicken auf den Eintrag Stored Process for Association Rules with Prompt wird der Stored Process mit einer Eingabeaufforderung ausgeführt, in der der Category Manager verschiedene Einstellungsmöglichkeiten erhält, um den Algorithmus für die Regelgenerierung zu steuern:



Bei Übernahme der Voreinstellungen erscheint folgende Ergebnistabelle, die für Weiterverarbeitung auch heruntergeladen werden könnte:



Endgültiges Fazit:

Es wurden mehrere Wege aufgezeigt, wie man Verbundkaufeffekte im Sortiment identifizieren kann. Man kann entweder hypothesenorientiert vorgehen und einen bestimmten Artikel als Ausgangsbasis festlegen (Kacheldiagramm-Darstellung) oder hypothesenfrei auf beliebige Kombinationen von Artikeln schauen (Heatmap-Darstellung). Wenn diese visuelle Analyse aufgrund der kombinatorischen Möglichkeiten von Artikeln zu mächtig wird, können die Daten über Mustererkennung in Form von Assoziationsregeln angereichert werden. Die Assoziationsregeln können entweder vorberechnet und an die Tabelle angespielt werden oder sie werden interaktiv über einen Stored Process erzeugt.

In allen Fällen käme der Category Manager aber sehr schnell zum Ergebnis, dass Holzkohlegrills eine hohe Verbundkauf-Wirkung zusammen mit dem Artikel Deutschlandfahne im Referenzzeitraum Mai 2012 entfaltet haben. Für eine neue Promotion zur aktuellen WM 2014 sollte man daher vielleicht wieder auf diese Artikel zurückgreifen, sofern man die Deutschlandfahne als Aktionsartikel geeignet bewerben möchte.

Das Beispiel zeigt darüber hinaus, wie man komplexe analytische Funktionalität in Visual Analytics so integrieren kann, dass ein Fachanwender in der Sortimentsplanung sie nutzen kann. Visual Analytics eignet sich als Analyse-Plattform im Rahmen der Bondaten-Analyse besonders, weil es ermöglicht, die gesamten Transaktionsdaten persistent in den Hauptspeicher zu laden und für interaktive Auswertungen und Analysen vorzuhalten. Als Fachanwender kommt man schneller zu fundierteren Entscheidungen, als über klassische Reporting-Werkzeuge.