

Hendrik Talkner

Business Intelligence mit Power BI - Band 2

ETL Prozesse, Datenmodellierung und Dashboarding für fortgeschrittene User

Wichtiger Hinweis

Alle Angaben in diesem Buch wurden vom Autor mit größter Sorgfalt erarbeitet und zusammengestellt. Trotzdem sind Fehler nicht ganz auszuschließen. Der Verlag und der Autor sehen sich deshalb gezwungen, darauf hinzuweisen, dass sie weder eine Garantie noch die juristische Verantwortung oder irgendeine Haftung für Folgen, die auf fehlerhafte Angaben zurückgehen, übernehmen können. Für die Mitteilung etwaiger Fehler sind Verlag und Autoren jederzeit dankbar.

Internetadressen oder Versionsnummern stellen den bei Redaktionsschluss verfügbaren Informationsstand dar. Verlag und Autor übernehmen keinerlei Verantwortung oder Haftung für Veränderungen, die sich aus nicht von Ihnen zu vertretenden Umständen ergeben.

Evtl. beigefügte oder zum Download angebotene Dateien und Informationen dienen ausschließlich der nicht gewerblichen Nutzung. Eine gewerbliche Nutzung ist nur mit Zustimmung des Lizenzinhabers möglich.

Inhaltsverzeichnis

1. VORWORT

- 1.1 ÜBER DIESES BUCH
- 1.2 AUFBAU DES BUCHES
- 1.3 ETL PROZESS
- 1.4 DATENMODELLIERUNG
- 1.5 REPORTING /DASHBOARDING

2. SZENARIO 1 - UMSATZAUSWERTUNG

- 2.1 EINLEITUNG
- 2.2 ETL PROZESS
 - 2.2.1 Einzelne Datei abfragen*
 - 2.2.2 Daten bereinigen*
 - 2.2.3 Zuletzt verwendete Quellen*
 - 2.2.4 Abfrageschritte*
- 2.3 DATENMODELLIERUNG UND REPORTING / DASHBOARDING
 - 2.3.1 Beziehungen zwischen Tabellen herstellen*
 - 2.3.2 Einfaches Measure erstellen, validieren und visualisieren*
 - 2.3.3 Mit Kalendertabellen arbeiten*
 - 2.3.4 Calculate() Funktion*
 - 2.3.5 All() Funktion in Kombination mit Calculate()*
 - 2.3.6 Calculate() in weiteren Variationen*
 - 2.3.7 AllSelected() in Kombination mit calculate()*

2.3.8 Formatierung des Dashboards

3. SZENARIO 2 - SOLL - IST VERGLEICH MIT KPIS

3.1 EINLEITUNG

3.2 ETL PROZESS

3.2.1 Dateien abfragen

3.2.2 Spalten erstellen

3.2.3 Spalten duplizieren

3.2.4 Benutzerdefinierte Spalte einfügen

3.2.5 Spalte aus Beispielen einfügen

3.2.6 Abfragen zusammenführen

3.2.7 Tabellen anfügen

3.3 DATENMODELLIERUNG UND REPORTING

3.3.1 GuV Erstellung mit Matrix Visualisierung

3.3.2 Measures mit switch() wechseln

3.3.3 GuV KPIs als Measure erstellen

3.3.4 Bedingte Formatierung (conditional formatting)

3.3.5 Bedingte Formatierung nach Farbskala

3.3.6 Bedingte Formatierung nach Regeln

3.3.7 Bedingte Formatierung nach Feldwert

3.3.8 Formatierung des Dashboards

4. SZENARIO 3 - PERSONALDASHBOARD

4.1 EINLEITUNG

4.2 ETL PROZESS

4.2.1 Daten entpivotieren

4.2.2 Tabellenstrukturen vertauschen & Ausfüllen

*4.2.3 Power Query Functions für
Massenabfragen nutzen*

4.2.4 Im M-Code arbeiten

4.2.5 Mehrere Abfragen anfügen

*4.2.6 Weitere Datenanreicherungen mit
benutzerdefinierten Spalten*

4.3 DATENMODELLIERUNG UND REPORTING

4.3.1 Visualisierungen erstellen und anpassen

*4.3.2 Anzahl der Arbeits, Krankheits und
Urlaubstage*

*4.3.3 Gesamtanzahl der geleisteten
Arbeitsstunden*

4.3.4 Durchschnittliche tägliche Arbeitszeit

4.3.5 Rankx() Funktion

*4.3.6 Formatierung und hinzufügen einer
weiteren Berichtsseite*

5. SCHLUSSWORT

1 Vorwort

1.1 Über dieses Buch

Liebe Leser,
ich freue mich sehr, dass Sie sich zum Kauf dieses Buches entschieden haben.

Da es in diesem Buch inhaltlich um eher fortgeschrittene Themen geht, gehe ich davon aus, dass Sie entweder mein bereits erschienenenes Buch „Business Intelligence mit Power BI – Der Einstieg in die Self Service BI Welt – Schritt für Schritt“ gelesen haben, oder aber im Umgang mit Power BI bereits über Grundwissen verfügen und dieses nun mit Hilfe dieses Buches ausbauen möchten.

In jedem Fall sollten Sie bereits wissen, um was es sich bei Power BI handelt und die grundlegende Funktionalität und Möglichkeiten kennen, weshalb ich mir im Gegensatz zu meinem ersten Buch ein umfangreiches allgemeines Einführungskapitel spare und hier direkt mit dem anwendbaren Inhalt starten werde.

Die Datenquellen, auf die im Laufe des Buches zugegriffen werden, können alle über eine Dropbox heruntergeladen werden. Die Anleitung und den Link hierzu finden Sie auf der letzten Seite dieses Buches unter dem Schlusswort. Ebenso finden Sie bei den Dateien die fertigen Power BI Desktop Dateien zum Download, die ich für dieses Buch erstellt habe (in welche Sie natürlich nur im Notfall einen Blick werfen müssen).

Sollten Sie Probleme mit dem Download haben, senden Sie mir nur eine kurze Mail an die unten aufgeführte Adresse und ich sende Ihnen die Dateien per Mail zu.

Ich empfehle, dieses Buch nicht einfach nur zu lesen, der Mehrwert für Sie wäre sicher nicht besonders groß. Sie sollten parallel zum Durchlesen dieses Buches die BI Umgebungen selbst nachzubauen und bei den etwas komplizierteren Stellen ein wenig über die Formeln nachdenken, bis Sie das Verhalten wirklich durchblickt haben. Ich denke, dass so die besten Lernfortschritte erzielt werden.

Und nun wünsche ich Ihnen viel Spaß beim Lesen und Durcharbeiten dieses Buches.

Bei Fragen, Anregungen, Hinweisen oder Ähnliches können Sie mich gerne unter folgender Mailadresse kontaktieren:

Info-pbi-buch@gmx.de

1.2 Aufbau des Buches

Beim Erstellen von Power BI Inhalten werden die folgenden Prozessschritte stets durchlaufen. Wobei es nicht zwangsläufig so ist, dass sie nacheinander durchlaufen werden müssen. Es ist völlig normal, zwischen den Schritten hin und her zu wechseln, wenn später festgestellt wird, dass am vorigen Schritt noch eine Änderung vorgenommen werden sollte.

1. Es werden Daten von einer Datenquelle angezapft und in eine brauchbare Form umgewandelt (ETL Prozess).
2. Es wird ein Datenmodell aufgebaut und Berechnungen werden durchgeführt. Man könnte auch sagen, es werden Fragen an die Daten gestellt, die man mit Berechnungslogiken beantwortet (Datenmodellierung).
3. Das Datenmodell wird visualisiert und den Reportkonsumenten zur Verfügung gestellt (Reporting, Dashboarding).

Diese drei Punkte werden Sie jedes Mal durchlaufen, wenn Sie mit Power BI arbeiten. Jeder einzelne Punkt kann dabei einfach zu lösen sein oder aber auch komplex und anspruchsvoll sein und auch sehr unterschiedlich von der Herangehensweise her.

Deshalb habe ich mir überlegt, für dieses Buch 3 für die Praxis gut anwendbare und brauchbare Szenarien vorzustellen und diese mit Ihnen zusammen aufbaue. Ein Szenario ist immer unterteilt in die Schritte „ETL Prozess“ und „Datenmodellierung & Reporting“. Die Prozessschritte Datenmodellierung und Reporting habe ich

zusammengefasst, da man doch sehr häufig zwischen den beiden Themen hin und herspringen muss.

Sowohl die Datenquellen als auch die fertigen Power BI Desktop Dateien stehen Ihnen in der oben angegebenen Dropbox zur Verfügung. Ich empfehle jedoch, die Datenmodelle selbst anhand des Buches aufzubauen und nur für den Fall, dass Ihnen etwas nicht gelingt, in die fertigen Dateien zu schauen.

Noch ein Hinweis zu den Screenshots. Mir ist bewusst, dass die Screenshots an der ein oder anderen Stelle etwas klein wirken. Es kommt mir jedoch nicht darauf an, dass man jedes Schriftzeichen auf dem Screenshot lesen kann, sondern darauf, dass Sie mit Hilfe meiner Pfeile und Rechtecke, die ich den Screenshots hinzufüge, schnell sehen können, wo auf Ihrem Bildschirm Sie ihren Fokus hinlenken sollten.

Da dieses Buch auf Fortgeschrittenen Niveau einzuordnen ist, werde ich beim Durchgehen der Szenarien nicht ständig auf die Basics eingehen. Es ist nicht einfach, hier einen guten Mittelweg zu finden, da es einerseits kein Zwang sein soll, mein erstes Buch zu kaufen und durchzuarbeiten, andererseits soll der Personenkreis, die mein erstes Buch gekauft und durchgearbeitet hat nicht das Gefühl bekommen, hier auf recycelte Inhalte zu stoßen. Deshalb wird es lediglich im ersten Szenario (neben vielen neuen Themen) ein paar wenige grundlegende Punkte geben, die bereits in meinem vorigen Buch behandelt wurden und im Schnelldurchlauf wiederholt werden.

Im Folgenden werde ich noch kurz etwas ausführlicher erläutern, was man unter den einzelnen 3 Prozessschritten (ETL, Datenmodellierung und Reporting / Dashboarding) versteht.

1.3 ETL Prozess

Die Abkürzung **ETL** steht ausgeschrieben für:

Extract

Transform

Load

Power BI enthält mit Power Query (Abfrageeditor) ein sehr mächtigste ETL Tool. Es ermöglicht Daten aus nahezu beliebiger Quelle anzuzapfen (extract), umzuformen (transform) und in das Power BI Datenmodell zu laden (load).

Power Query ist meiner Ansicht nach so mächtig, da man einen komplexen ETL Prozess, der bisher von Programmierprofis mit beispielsweise VBA und SQL Lösungen konzipiert und programmiert wurde, nun per Mausklick und ohne Programmierkenntnisse selbst umsetzen kann.

Dies ist nicht immer trivial und Datenbank und Programmierkenntnisse sind trotzdem hilfreich und führen dazu, schneller an die gewünschte Lösung des Problems zu kommen, aber dennoch sind sie in den meisten Fällen nicht notwendig. Hinter Power Query steckt die Programmiersprache „M“.

Extract:

Es besteht die Möglichkeit, in einem Datenmodell unterschiedliche Quelldaten anzuzapfen, also zum Beispiel Exceldateien, Textdateien, PDF Dateien, SQL Datenbanken, Internetseiten, Sharepoints usw. und diese in einem Datenmodell zu vereinen. Die Schnittstelle zu den einzelnen Datenquellentypen stehen in Power Query zur Verfügung und müssen lediglich ausgewählt werden.

Transform:

In diesem Schritt werden die angezapften Daten weiterverarbeitet, um Sie für die spätere Verwendung im Datenmodell brauchbar zu machen. Dies kann zum Beispiel das Filtern von Daten umfassen, das Entfernen von Spalten, Änderung von Datentypen, Zusammenführen von Abfragen oder das Gruppieren von Datensätzen. Ebenso kann man

hierzu auch das Ersetzen von Werten oder das Hinzufügen von Spalten zählen. Auf alle diese Möglichkeiten werden wir im Laufe des Buches detailliert eingehen.

Load:

Nachdem die Daten angezapft und für unsere Zwecke transformiert wurden, werden Sie in das Power BI Datenmodell geladen und können dort weiterverarbeitet werden. Dies ist im Falle von Power Query in Verbindung mit Power BI recht unkompliziert, da das Laden der Daten einfach per Knopfdruck geschieht.

1.4 Datenmodellierung

An dieser Stelle verfügt das Datenmodell innerhalb von Power BI über wahrscheinlich mehrere für sich stehende Tabellen aus unterschiedlichen Datenquellen, die losgelöst voneinander Daten beinhalten. Im Sinne der Power BI Datenmodellierung werden diese Tabellen, dort wo es Sinn macht miteinander verknüpft, um komplexere, weiterführende Berechnungen mit DAX zu ermöglichen. Die Berechnungen werden in berechneten Spalten, Tabellen und Measures durchgeführt. Ich rede gerne davon, dass hier Fragen an die Daten gestellt werden. Zum Beispiel hat man eine Tabelle, die Datensätze enthält, aus denen Umsätze nach Kunden hervorgehen. In einer weiteren Tabelle befinden sich Datensätze, aus denen hervorgeht, welcher Vertriebsmitarbeiter für welchen Kunden verantwortlich ist. Hier könnte man nun die Frage an die Daten richten: „Für wie viel % des gesamten Umsatzes im Jahr 2018 war Herr Müller verantwortlich?“. Um die Frage zu beantworten, müssen die einzelnen Tabellen miteinander in Beziehung gesetzt werden und ein Measure wird mit einer Berechnungslogik versehen, die genau die Frage beantwortet, die gestellt wurde.

1.5 Reporting /Dashboarding

Hierbei geht es darum, die durchgeführten Berechnungen und das erstellte Datenmodell so zu visualisieren und für den Reportkonsumenten einfach verständlich und mehrwertbringend aufzubereiten. Ein fertiger Report soll möglichst interaktiv gestaltet sein und intuitiv zu bedienen sein und die Fragen, die der Konsument an die Daten stellen würde, beantworten können. Auch die Frage nach der Aktualisierung der Berichte und deren Bereitstellung und Zugriffsberechtigungen zähle ich zu diesen Punkt. Hochladen in Power BI Online, Aktualisierung und Berechtigungen werden in diesem Buch jedoch nicht thematisiert, dieses Thema habe ich abschließend in meinem ersten Buch behandelt. Ein Szenario endet in diesem Buch stets an der Stelle, an der der Bericht fertig gestellt ist und zum Hochladen in Power BI Online bereit ist.

2 Szenario 1 - Umsatzauswertung

2.1 Einleitung

Das erste Szenario dient dem Einstieg. Hier bauen wir nochmal eine ähnliche Umgebung wie in meinem zuvor erschienenen Buch auf. Im direkten Vergleich werden wir jedoch ein paar Punkte auf einfachere Art und Weise lösen und zudem bereits ein paar weiterführende Techniken anwenden.

Ziel dieses Szenarios ist es, aus der uns vorliegenden Datenquelle ein professionell aufbereitetes Umsatzdashboard zu kreieren, aus welchem hervorgeht, in welchen Ländern sich welche Produkte des Unternehmens wie gut verkaufen.

Als Datenquelle dient das öffentlich zugängliche Northwind Dataset, welches Sie im Ordner Szenario 1 der Dropbox finden. Bei der Datenquelle handelt es sich um eine Exceldatei mit mehreren Tabellenreiter. Laden Sie am besten jetzt die Datenquellen auf Ihren PC lokal herunter (falls noch nicht geschehen). Ist dies geschehen erstellen Sie bitte eine neue Power BI Desktop Datei für das erste Szenario.

Hinweis zu den herunterladbaren Power BI Desktop Dateien. In den Power BI Dateien sind Verbindungen zu den Datengrundlagen hinterlegt. Im Fall der Beispieldateien handelt es sich um Exceldateien, die auf meinem persönlichen Rechner gespeichert sind. Damit Sie vernünftig mit den Power BI Dateien arbeiten können, ist es notwendig, den Pfad dieser Datengrundlagen anzupassen auf den Pfad, wo Sie die Dateien auf Ihrem Rechner gespeichert haben.

Dies ist sehr einfach. Nachdem Sie eine Power BI Datei geöffnet haben, klicken Sie im Menüband „Home“ auf „Abfragen bearbeiten“. Dadurch gelangen Sie in Power Query.



In Power Query angelangt klicken Sie im Menüband „Home“ auf „Datenquelleneinstellungen“.



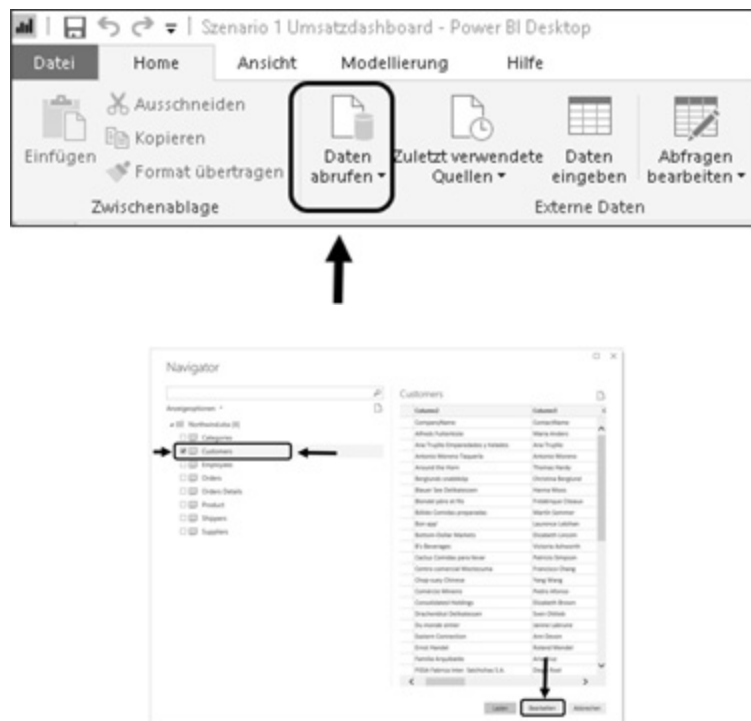
Im Fenster, welches als nächstes auftaucht, sehen Sie in der Mitte alle Datenquellen aufgelistet, die diese Power BI Datei enthält. Sie müssen lediglich unten links auf „Quelle ändern“ klicken und dort den Pfad auswählen, wo Sie die entsprechende Datei abgelegt haben, mit „OK“ bestätigen und dann „Schließen“ betätigen.



2.2 ETL Prozess

2.2.1 Einzelne Datei abfragen

Als Datengrundlage haben wir in diesem Fall eine Exceldatei mit mehreren Tabellenreiter. Wir werden für jeden Tabellenreiter eine separate Abfrage erstellen und die Daten nach unseren Bedürfnissen bereinigen und anreichern. Öffnen Sie dafür bitte in Power BI den Abfrageeditor (Power Query) indem Sie auf „Daten abrufen“ gehen und im Anschluss „Excel“ auswählen.



Sie gelangen in eine Dateibrowser Ansicht und können nun die bei Ihnen lokal abgespeicherte Exceldatei suchen, auswählen und mit „Öffnen“ bestätigen. Im nun erscheinenden Fenster besteht die Möglichkeit, auszuwählen, welchen Tabellenreiter der Exceldatei wir