

Kapitel 26

Daten importieren und aufbereiten: Mit modernem Werkzeug schneller zum Ziel

In diesem Kapitel:

Einstieg in Power Query	799
CSV-Dateien fehlerfrei importieren und aufbereiten	800
Listen abgleichen ganz ohne Formeln	808
Tabellen von Internetseiten importieren und in eigenen Analysen nutzen	813
Duplikate finden und löschen, auch über mehrere Tabellen hinweg	826

Einstieg in Power Query

Nicht immer liegen die Daten für Auswertungen in der passenden Form vor und eine zeitintensive Datenaufbereitung ist notwendig. In Excel 2016 erledigen Sie solche Arbeiten zeitsparend, wenn Sie das moderne Werkzeug nutzen, das unter dem Namen Power Query bekannt wurde.

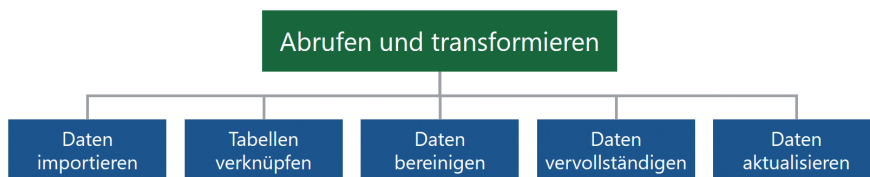


Abbildung 26.1: Wichtige Einsatzmöglichkeiten der Befehle aus der Gruppe Abrufen und transformieren

Hinweis Im Unterschied zu früheren Excel-Versionen ist Power Query jetzt fester Bestandteil von Excel und nicht ein Add-In, das erst zusätzlich installiert und aktiviert werden muss. Allerdings heißt es nicht mehr Power Query, sondern wird auf der Registerkarte *Daten* unter *Abrufen und transformieren* geführt (Abbildung 26.2).



Abbildung 26.2: Hier finden Sie die Befehle zum Importieren und Aufbereiten von Daten

Über die Gruppe *Abrufen und transformieren* können Sie Daten

- aus unterschiedlichsten Quellen in ein Arbeitsblatt oder in das Datenmodell importieren,
- ohne Formeln miteinander verknüpfen,
- bereinigen und in eine für Auswertungen geeignete Form bringen,
- mit eigenen Berechnungen vervollständigen,
- per Mausklick oder automatisch aktualisieren – ganz ohne VBA.

Mit einem Satz: Zeitintensive und lästige Routinearbeiten beim Importieren und Aufbereiten Ihrer Daten gehören der Vergangenheit an.

CSV-Dateien fehlerfrei importieren und aufbereiten

Viele Programme von Drittanbietern exportieren die Daten im CSV-Format. Beim Import in Excel sorgen länderspezifische Datums- und Zahlenformate dann regelmäßig für Probleme. Aufwendige Nacharbeiten sind die Folge.

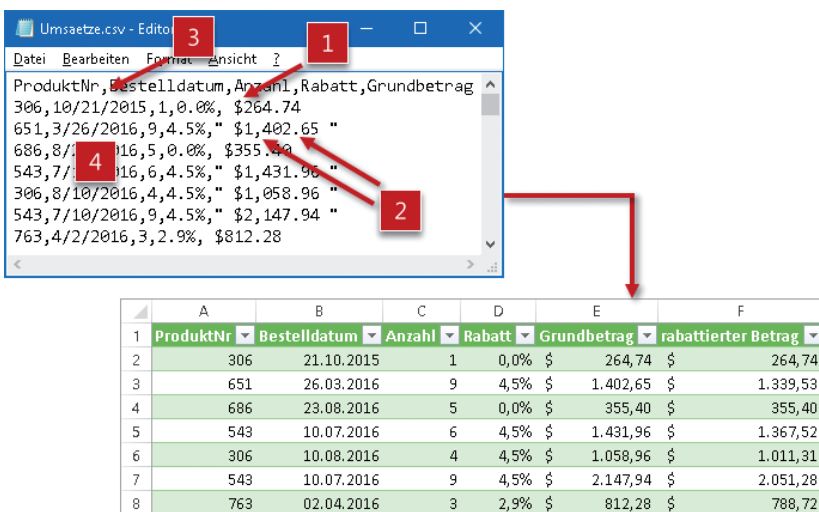


Abbildung 26.3: Die CSV-Datei und das Ergebnis des Imports mit zusätzlicher Spalte in Excel

Beim Import der in Abbildung 26.3 dargestellten CSV-Datei in Excel sind die folgenden vier Importprobleme zu meistern, die Tabelle 26.1 im Detail aufführt:

	Vorher	Nachher
1	Vor jeder Zahl befindet sich ein Dollarzeichen. Allerdings wird es beim Import nicht als Währungssymbol erkannt. Folglich würden alle Beträge als Text interpretiert werden.	In der Zelle steht nur der Betrag. Das Dollarzeichen wird über ein Zahlenformat angezeigt.
2	Dezimal- und Tausendertrennzeichen entsprechen dem US-Standard.	Komma und Punkt werden jetzt nach deutschem Standard angezeigt.
3	Als Listentrennzeichen wird das Komma verwendet. Da es auch zur Zifferngruppierung dient, stehen alle Beträge > 1000 Dollar in Anführungszeichen.	Trotz abweichendem Listentrennzeichen (Komma statt Semikolon) werden die Daten richtig auf die Spalten verteilt. Beträge > 1000 Dollar werden korrekt als Zahlen angezeigt.
4	Das Datum folgt dem Schema Monat/Tag/Jahr.	Das Datum wird nach deutschem Standard angezeigt (TT.MM.JJJJ).

Tabelle 26.1: Mit Power Query lösen Sie typische Importprobleme

Mit den Befehlen der neuen Gruppe *Abrufen und transformieren* ist das kein Problem mehr. Sie können die Daten

- bereits beim Import passend aufbereiten,
- um eigene Berechnungen ergänzen,
- das Gebietsschemaproblem lösen, damit zum Beispiel Währungsbeträge und Trennzeichen korrekt erkannt werden, und
- die Datenaufbereitung automatisieren – ganz ohne VBA.

Online Sie sitzen am Computer und möchten gleich mitmachen? Dann öffnen Sie eine leere Arbeitsmappe mit **[Strg]+[N]**.

Die CSV-Datei in den Abfrage-Editor laden

Laden Sie zuerst die Daten in den neuen Abfrage-Editor. Dort stehen Ihnen anschließend viele nützliche Werkzeuge zur Bereinigung und Aufbereitung zur Verfügung:

1. Wählen Sie über *Daten/Abrufen und transformieren/Neue Abfrage/Aus Datei* den Befehl *Aus CSV*.
2. Navigieren Sie im Dialogfeld *Daten importieren* zu der Datei *Kap_26_Umsaetze.csv* und klicken Sie auf *Importieren*.
3. Im nächsten Dialogfeld sehen Sie das Ergebnis der Analyse der CSV-Datei. Power Query hat das Komma als Spaltentrennzeichen korrekt erkannt. Ein Klick auf *Bearbeiten* lädt die Datei in den Editor.

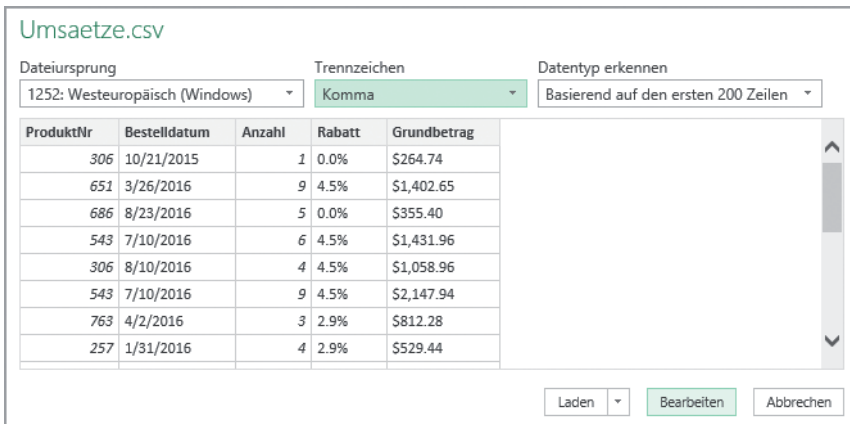


Abbildung 26.4: Power Query hat in der CSV-Datei das Komma als Spaltentrennzeichen automatisch erkannt

Die Vorzüge des Abfrage-Editors kennenlernen

Machen Sie sich zuerst mit dem Abfrage-Editor vertraut (siehe dazu Abbildung 26.5):

- Im Editor sehen Sie links eine Vorschau der CSV-Datei [1].
- Rechts oben im Aufgabenbereich *Abfrageeinstellungen* [2] werden die Eigenschaften aufgelistet.
- Darunter wird jeder Arbeitsschritt protokolliert [3], den Sie beim Import oder bei der Datenaufbereitung durchführen.
- Anhand dieses Schritt-für-Schritt-Protokolls können Sie bei Bedarf einzelne Schritte rückgängig machen. Bewegen Sie dazu die Maus auf den betreffenden Eintrag und klicken Sie auf das links eingblendete Lösch-Kreuz [4].
- Die Reihenfolge der Schritte ändern Sie, indem Sie diese mit gedrückter linker Maustaste an die gewünschte Stelle verschieben.

Hinweis Wenn der Aufgabenbereich *Abfrageeinstellungen* am rechten Bildschirmrand nicht angezeigt wird, blenden Sie ihn über *Ansicht/Abfrageeinstellungen* wieder ein.

Hinweis Hinter jedem Abfrageschritt verbirgt sich eine Power Query-Formel, die Sie nur bei eingblendeter Bearbeitungsleiste sehen können. Aktivieren Sie hierzu das gleichnamige Kontrollkästchen auf der Registerkarte *Ansicht*.

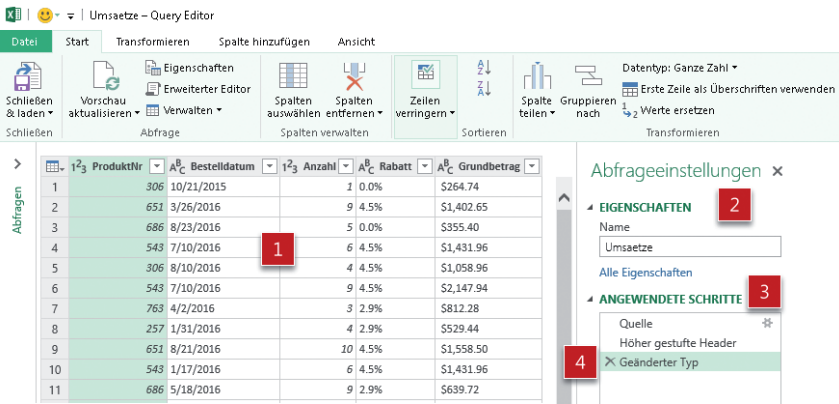


Abbildung 26.5: Datenanzeige und Protokoll im Abfrage-Editor von Power Query

Den Datentyp einzelner Spalten ändern

Beim Importieren von CSV-Dateien versucht Power Query, die Namen und Datentypen der Spalten automatisch zu erkennen. Daher enthält die Liste *Angewendete Schritte* bereits die Einträge *Höher gestufte Header* und *Geänderter Typ*.

An dem Symbol neben der Spaltenüberschrift erkennen Sie den Datentyp: Die Spalten *ProduktNr* und *Anzahl* hat Power Query richtig als *Ganze Zahl* erkannt, während die übrigen drei Spalten noch vom Typ *Text* sind. Dieses Manko korrigieren Sie mit wenigen Mausklicks.

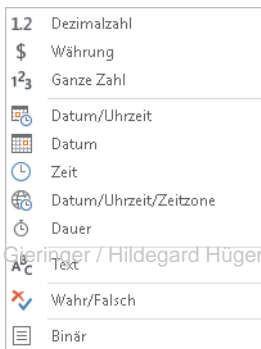


Abbildung 26.6: Übersicht der Datentypen von Power Query

Hinweis

Ob Power Query Zahlen und Datumsangaben aus der CSV-Datei korrekt interpretiert, hängt mit den Einstellungen in der Windows-Systemsteuerung zusammen. Stimmen die Einstellungen für Datumsangaben, Dezimal- und Tausendertrennzeichen mit den in der CSV-Datei verwendeten Zeichen überein, erkennt Power Query die Datentypen korrekt. Entsprechen die Windows-Einstellungen dem deutschen Standard, werden im Beispiel weder das Bestelldatum noch die Dezimalzahlen korrekt erkannt, denn sie folgen dem US-amerikanischen Standard.

Beim Transformieren das Gebietschema korrigieren

Das Bestelldatum folgt dem in den USA gängigen Schema Monat/Tag/Jahr und nicht dem in Deutschland verwendeten Schema TT.MM.JJJJ. Damit Excel später das Datum korrekt interpretiert, gehen Sie wie folgt vor:

1. Per Rechtsklick auf die Spaltenüberschrift *Bestelldatum* wählen Sie *Typ ändern/Mit Gebietsschema*.
2. Als Datentyp wählen Sie *Datum* und als Gebietsschema *Englisch (USA)*. Bestätigen Sie mit *OK*.
3. Ändern Sie analog dazu den Datentyp der Spalte *Grundbetrag* auf *Währung*.

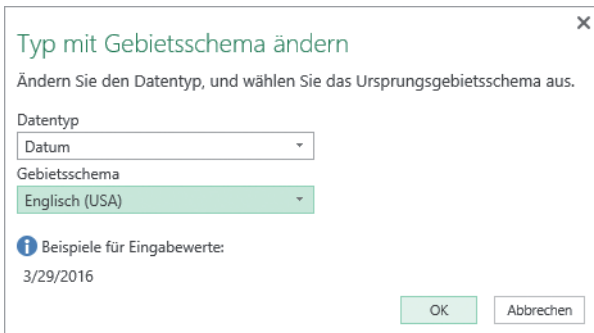


Abbildung 26.7: So wird der Import von länderspezifischen Datums- und Zahlenformaten zum Kinderspiel

Die Spalte »Rabatt« in einen Zahlenwert verwandeln

Der Rabatt lässt sich nicht direkt in eine Dezimalzahl verwandeln. Schuld ist das Prozentzeichen. Nachdem Sie das störende Prozentzeichen entfernt haben, wandeln Sie die Werte der Spalte in Dezimalzahlen um. Anschließend dividieren Sie alle Werte der Spalte durch 100, um den Rabatt wieder in einen Prozentwert zu ändern. Gehen Sie dazu wie folgt vor:

Das Prozent-Zeichen entfernen

1. Wählen Sie per Rechtsklick auf die Spaltenüberschrift *Rabatt* den Befehl *Werte ersetzen*.
2. Bei *Zu suchender Wert* geben Sie ein %-Zeichen ein. Lassen Sie *Ersetzen durch* leer und klicken Sie auf *OK*.

Den Rabatt in eine Dezimalzahl verwandeln

Noch interpretiert Power Query die Spalte *Rabatt* als Textspalte. So machen Sie daraus eine Dezimalzahl:

1. Per Rechtsklick auf die Spaltenüberschrift *Rabatt* wählen Sie *Typ ändern/Mit Gebietsschema*.
2. Als Datentyp wählen Sie *Dezimalzahl* und als Gebietsschema *Englisch (USA)*. Bestätigen Sie mit *OK*.

Den Rabatt wieder in einen Prozentwert ändern

Durch das Entfernen des Prozentzeichens ist die Information verloren gegangen, dass es sich bei den Zahlen um Prozentwerte handelt. Aus 4,5% (also 0,045) wurde 4,5. Somit sind alle Rabattwerte noch um das Hundertfache zu groß. Jeder Wert der Spalte *Rabatt* muss also noch durch 100 dividiert werden. Auch dies geht ganz bequem ohne Formel:

1. Markieren Sie per Mausclick die Spalte *Rabatt*.
2. Klicken Sie auf der Registerkarte *Transformieren* in der Gruppe *Zahlenspalte* auf *Standard/Dividieren* und geben Sie im Dialogfeld 100 ein. Schließen Sie das Dialogfeld über *OK*.

Tip

Die Datentypen der Spalten lassen sich mit einem Blick anhand der Symbole in den Spaltenüberschriften erkennen. Eine Übersicht aller Symbole finden Sie in Abbildung 26.6. Die Ausrichtung innerhalb der Zelle lässt ebenfalls Rückschlüsse auf den Datentyp zu: Zahlen und Datumsangaben sind wie in Excel rechtsbündig, Texte linksbündig.

Die fertige Abfrage ins Arbeitsblatt laden

Nachdem Sie mit der Aufbereitung der Daten fertig sind, benennen Sie die Abfrage um und laden das Ergebnis ins Arbeitsblatt:

1. Benennen Sie die Abfrage in den *Abfrageeinstellungen* am rechten Bildschirmrand in *Umsätze* um.

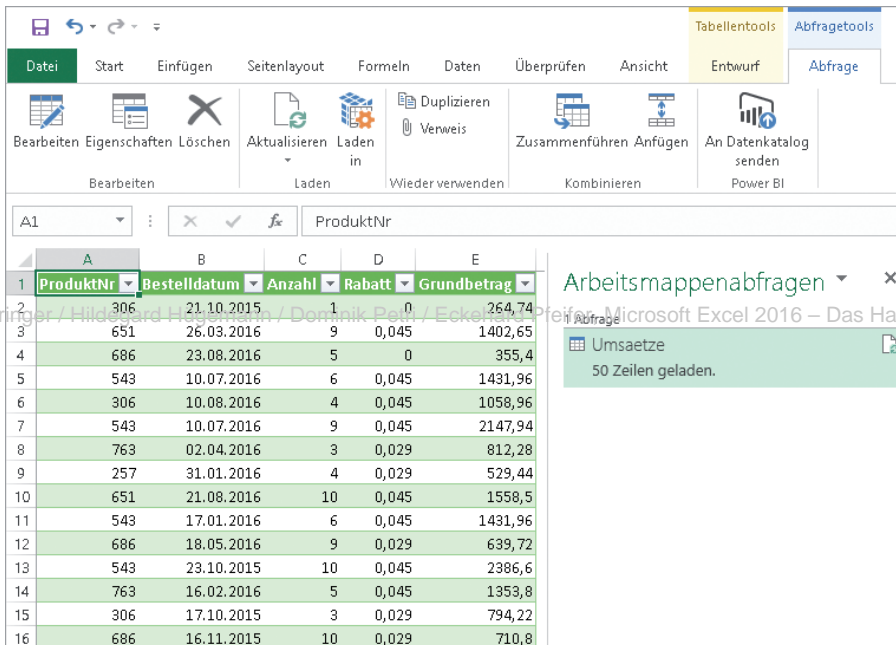


Abbildung 26.8: Die importierte CSV-Datei als »intelligente« Tabelle in Excel

2. Starten Sie den Import über *Start/Schließen & Laden*. Dadurch wird der Arbeitsmappe ein neues Tabellenblatt hinzugefügt. Darauf befinden sich die importierten Daten in einer »intelligenten« Tabelle, die den gleichen Namen wie die Abfrage trägt. Im Aufgabenbereich *Arbeitsmappenabfragen* sehen Sie, dass insgesamt 50 Zeilen geladen wurden.

Hinweis Wird der Aufgabenbereich *Arbeitsmappenabfragen* nicht angezeigt, blenden Sie ihn bei Bedarf über *Daten/Abfragen und transformieren/Abfragen anzeigen* wieder ein.

Die Tabelle um eigene Berechnungen ergänzen

Nach dem Import möchten Sie die Daten in Excel weiter analysieren und auswerten. Dazu benötigen Sie den rabattierten Betrag, also den Grundbetrag abzüglich des gewährten Rabatts. Dazu ergänzen Sie die Tabelle um zwei weitere Spalten: In der ersten ermitteln Sie den Rabattbetrag und ziehen diesen in der zweiten vom Grundbetrag ab.

Natürlich könnten Sie dies mit Excel-Formeln im Arbeitsblatt erledigen, doch bequemer und ganz ohne Formeleingabe geht's wiederum mit Power Query:

Die bestehende Abfrage bearbeiten

1. Markieren Sie eine Zelle in der »intelligenten« Tabelle. Dadurch wird die Registerkarte *Abfragetools/Abfrage* angezeigt.
2. Öffnen Sie den Abfrage-Editor per Klick auf *Abfragetools/Abfrage/Bearbeiten/Bearbeiten*.

Zwei Spalten miteinander multiplizieren

Ermitteln Sie den Rabattbetrag durch Multiplikation des Rabattsatzes mit dem Grundbetrag:

1. Markieren Sie nacheinander mit gedrückter **[Strg]**-Taste die Spalten *Rabatt* und *Grundbetrag*.
2. Klicken Sie auf der Registerkarte *Spalte hinzufügen* in der Gruppe *Aus Zahl auf Standard/Multiplizieren*.
3. Benennen Sie die neue Spalte per Doppelklick auf den Spaltenkopf in *Rabattbetrag* um.

Eine Spalte von einer anderen Spalte subtrahieren

Ziehen Sie den Rabatt- vom Grundbetrag ab, erhalten Sie den rabattierten Betrag:

1. Markieren Sie nacheinander mit gedrückter **[Strg]**-Taste die Spalten *Grundbetrag* und *Rabattbetrag*. Wichtig: Nur bei dieser Klickreihenfolge werden im nun folgenden Schritt die Rabattbeträge von den Grundbeträgen abgezogen.
2. Klicken Sie auf der Registerkarte *Spalte hinzufügen* in der Gruppe *Aus Zahl auf Standard/Subtrahieren*.
3. Benennen Sie die neue Spalte per Doppelklick auf den Spaltenkopf in *rabattierter Betrag* um.

Überflüssige Spalten löschen und die geänderte Abfrage ins Tabellenblatt laden

1. Die Spalte mit den Rabattbeträgen benötigen Sie nun nicht mehr. Wählen Sie per Rechtsklick auf die Spaltenüberschrift *Rabattbetrag* den Befehl *Entfernen*.
2. Klicken Sie auf *Start/Schließen & Laden*, um die geänderte Abfrage ins Tabellenblatt zu laden.

Mit Zahlenformaten für den Feinschliff sorgen

Der Grundbetrag in der CSV-Datei enthielt ein vorangestelltes Dollarzeichen. Daher hat Power Query die Beträge zuerst nicht als Zahl, sondern als Text interpretiert. Durch die Umwandlung in eine Dezimalzahl ist die Währungsinformation verloren gegangen. Außerdem wird der Rabattnsatz noch nicht als Prozentzahl angezeigt. Beides korrigieren Sie über entsprechende Zahlenformate in Excel:

1. Markieren Sie eine Zahl in der Spalte *Rabatt* und drücken Sie anschließend die Tastenkombination `[Strg]+[Leertaste]`, um alle Zahlen zu markieren.
2. Klicken Sie dann auf der Registerkarte *Start* in der Gruppe *Zahl* auf die Schaltfläche *Prozentformat* oder benutzen Sie die Tastenkombination `[Strg]+[⇩]+[%]`. Die Rabattnsätze werden ohne Nachkommastellen und mit einem angehängten Prozentzeichen formatiert angezeigt.
3. Klicken Sie in der Gruppe *Zahl* auf die Schaltfläche *Dezimalstelle hinzufügen*, damit die Prozentwerte eine Nachkommastelle erhalten.
4. Markieren Sie analog zu Schritt 1 die Spalten *Grundbetrag* und *rabattierter Betrag*.
5. Klicken Sie in der Gruppe *Zahl* rechts neben dem Symbol mit dem Münzstapel auf den kleinen Pfeil und wählen Sie *\$ Englisch (Vereinigte Staaten)*. Die Beträge werden damit in US-Dollar angezeigt.



Abbildung 26.9: Mit diesem Buchhaltungszahlenformat ergänzen Sie die Zahlen um das Dollarzeichen

/ Dietmar Gieringer / Hildegard Hügemann / Dominik Petri / Eckehard Pfeifer, Microsoft Excel 2016 – Das Handbuch, O'Reilly, ISBN

Hinweis

Mehr zu Zahlenformaten in Excel finden Sie in den Kapiteln 6 und 7.

Fazit: Der Import mit Power Query spart Zeit und vermeidet Fehler

Anhand dieses Beispiels haben Sie gesehen, wie einfach der Import und das gleichzeitige Aufbereiten von Daten in Excel 2016 geworden ist. Importprobleme durch länderspezifische Datums- und Zahlenformate sowie aufwendige Nacharbeiten gehören der Vergangenheit an. Der neue Datenimport hat unter anderem folgende Vorteile: