

Business Intelligence

Neue Tools für Datenimport und -analyse in Excel

Von Stephan Blum, Dipl.-Finanzwirt (FH), Betriebswirt (VWA), Immobilienbewerter (IHK), Staatlich geprüfter Informatiker



Normalerweise sind BI-Tools teuer – doch Microsoft Excel kann in vielen Bereichen als kostengünstige Alternative dienen.

MS Excel ist weit mehr als „nur“ eine Tabellenkalkulation. Insbesondere die neueren Excel-Versionen 2013 und 2016 bieten Möglichkeiten zur Aufbereitung und Analyse von Massendaten, die bislang teuren und komplexen Speziallösungen vorbehalten waren. Da diese neuen Features von Excel noch weithin wenig bekannt sind, wird vorliegend ein Überblick über den Themenbereich gegeben, der die Leser darauf aufmerksam machen will, dass es sich lohnt, sich die Sache einmal näher anzuschauen.

1. Was ist Business Intelligence?

Hinter dem Schlagwort Business Intelligence (BI) verbirgt sich nichts anderes als die Analyse von großen Datenbeständen aus dem wirtschaftlichen Bereich. Ein Beispiel für eine BI-Anwendung ist die Prüfung der Finanzbuchhaltung durch die Betriebsprüfer mittels elektronischem Datenzugriff und dem Programm IDEA (etwas Ähnliches bietet Datev unter dem Namen ACL an).

Eine Liste mit hunderttausenden von Buchungssätzen ist für sich genommen völlig aussageelos. Erst durch die Zusammenfassung gleichartiger Daten in verschiedenen Sichten geben diese ihre Geheimnisse preis. So kann der

Betriebsprüfer beispielsweise die täglichen Kassenbuchungen kumulieren und auf diese Weise für jeden einzelnen Tag herausfinden, ob Kassenfehlbeträge vorhanden waren. Mit „bloßem Auge“ wäre eine solche Auswertung über mehrere Jahre nicht möglich – und aufgrund des Zeitaufwands auch nicht mit Papier und Taschenrechner.

BI-Tools bieten dabei verschiedene Analysemöglichkeiten, wie zum Beispiel Pivottabellen, die man sich als Datenbankabfragen in Form von Kreuztabellen vorstellen kann (mehr dazu weiter unten). Außerdem lassen sich Kennzahlen bilden und grafisch auswerten.¹ Ebenso können die Daten um eigene Formeln und Berechnungen ergänzt werden.

Selbstverständlich beschränken sich die Anwendungsmöglichkeiten von BI nicht auf die Betriebsprüfung. Alle Unternehmen müssen große Datenbestände analysieren und auswerten, zumindest dann, wenn man ein sinnvolles Controlling betreiben und die Entwicklung des Betriebs nicht dem Zufall überlassen möchte.

¹ Sogenannte KPIs oder Key Performance Indicators.

Man stelle sich beispielsweise ein Unternehmen vor, welches zum Vertrieb seiner Produkte mehrere Außendienstmitarbeiter beschäftigt. Angenommen, die Fakturierung erfasst zu jedem Umsatz auch den jeweiligen Vertreter. Dann ergibt sich während eines Jahres eine Liste mit vielen tausend Einträgen, von denen jeder für einen einzelnen Umsatz steht und die in ausgedruckter Form oder als PDF-Datei keine weitergehenden Analysen zulassen würde.

Importiert man diese Liste nun aber in eine BI-Software, dann lassen sich „auf Mausklick“ die verschiedensten Auswertungen erstellen und grafisch sowie in Tabellenform oder auch in Landkarten (sofern die Daten einen geografischen Bezug haben) darstellen. Aus den unzähligen Fragestellungen, die man auf diese Weise beantworten kann, nachfolgend einige wenige, die die Möglichkeiten der Datenanalyse verdeutlichen:

- Wie groß waren die Umsätze in den einzelnen Regionen?
- In welchen Monaten wurden die höchsten Umsätze erzielt?
- Welche Vertreter waren am erfolgreichsten?
- Welches waren die erfolgreichsten Produkte?

Es sind auch alle möglichen Kombinationen aus den oben genannten (und vielen weiteren) Fragestellungen denkbar, etwa:

- Welcher Vertreter hat am meisten Verträge über Produkt A abgeschlossen?
- Gibt es jahreszeitliche Unterschiede, die von Region zu Region variieren?

Mit wenigen Formeln kann man Kennzahlen bilden und in Ampelform darstellen. Mit ihrer Hilfe lässt sich die Entwicklung bestimmter Größen, zum Beispiele Vergleiche zum Vorjahreszeitraum oder zwischen Regionen, Produkten oder Vertretern, ablesen.

Generell stellt sich in der Steuerberatungspraxis (neben Auswertungen der genannten Art in eigener Sache) die Notwendigkeit, einerseits im Rahmen der betriebswirtschaftlichen Beratung die bestmögliche Entscheidungsgrundlage auf Basis der Unternehmensdaten zu entwickeln, und andererseits auch bei Betriebsprüfungen (oder noch besser in deren Vorfeld) mögliche Schwachpunkte des Zahlenwerks aufzuspüren. Hierdurch kann man sich auf die zu erwartenden Prüfungsfeststellungen einstellen. Es ist im Interesse der Steuerberater, möglichst Waffengleichheit im Bereich der EDV-Analyse herzustellen. Darunter fällt auch die Auseinandersetzung mit den statistischen Prüfverfahren („Benford's Law“ und Chi-Quadrat-Test), die mit einer BI-Software (und auch direkt in Excel) problemlos nachvollzogen werden können.

2. Die BI-Tools von Microsoft Office

Bislang waren die Hürden für den Einstieg in BI sehr hoch. Spezialsoftware aus diesem Bereich ist sehr teuer und liegt kaum im EDV-Budget einer mittleren Kanzlei. Zudem ist der Einarbeitungsaufwand in die Speziallösungen sehr groß und wenigen Datenanalysten vorbehalten. Es ist kaum möglich, sich mit derartigen Lösungen nur nebenher, also neben dem laufenden Tagesgeschäft, zu beschäftigen.

Inzwischen stellt sich die Situation allerdings völlig anders dar: Die neuen Excel-Versionen 2013 und 2016 enthalten ein Zusatzprogramm namens „Power Pivot.“²

In Verbindung mit dem – ebenfalls in Excel enthaltenen – Tool „PowerQuery“ ist es ein Leichtes geworden, Daten aus den verschiedensten Quellen in Excel zu importieren und zu analysieren. Für den normalen alltäglichen Gebrauch können diese neuen BI-Tools in MS Office die teuren Spezialprogramme durchaus ersetzen. Und nicht nur das, die Bedienung ist überdies relativ einfach, sodass die Grundlagen bereits in einem einzigen Tagesseminar vermittelt werden können³.

2.1. PowerQuery

PowerQuery ist ein in Excel enthaltenes Tool zum Einlesen und zur Aufbereitung von Daten⁴. Die Informatiker sprechen hier von einem „ETL-Tool“. ETL steht für „Extract, Transform & Load“, also „Extrahieren, Transformieren & Laden“. Mit „Extrahieren“ ist das „Herausziehen“ der Daten aus – nahezu – beliebigen Vorkommnissen gemeint. Das können Datenbanken aller Art und Herkunft sein, wie z. B. Datev-Daten, Exportdateien aus elektronischen Registrierkassen, Access-Datenbanken oder auch Excel-Dateien.

Der Import und die Aufbereitung der Daten für die weitere Analyse sind in der Regel die zeitaufwendigste und schwierigste Aufgabe des Datenanalysten. Sobald die Daten einmal in einer verwertbaren Form vorliegen, ist der Rest fast schon ein Kinderspiel. Entsprechend hilfreich ist PowerQuery, da es sehr mächtige Werkzeuge für den Datenimport und die Aufbereitung bietet und überdies leicht erlern- und bedienbar ist.

Wichtig ist lediglich, dass die Daten eine Datenbankstruktur aufweisen, also in Tabellen angeordnet sind, bei denen jeweils in derselben Spalte Informationen derselben Art und vom selben Datentyp (z. B. Texte oder Geldbeträge) stehen (Datenbank-Felder), während die einzelnen Zeilen inhaltlich verschiedene Informationen enthalten, die jedoch zu einer Einheit gehören (die einzelnen Zellen sind die „Datensätze“). Die Spalten haben jeweils Überschriften, die sogenannten Feldbezeichnungen.

² Die erste Excel-Version, für die Power Pivot erhältlich war, war Excel 2010 – allerdings war diese Version noch nicht auf dem hohen Niveau von Excel 2013 und 2016.

³ Neben den in Excel integrierten BI-Tools gibt es von Microsoft auch eine Version, die von Excel unabhängig und für sich alleine lauffähig ist („Power BI Desktop“). Auf dieses Produkt wird im Rahmen des vorliegenden Beitrags jedoch nicht eingegangen.

⁴ In Excel 2013 ist PowerQuery unter einem eigenen Register gleichen Namens zu finden; in Excel 2016 wurde die Funktionalität von PowerQuery direkt in das Register „Daten“ integriert. Es ist also nicht verschwunden, sondern hat nur kein eigenes Register mehr.

Ein Beispiel für eine solche Datenbankstruktur könnte eine Kundenliste wie in Abbildung 1 dargestellt sein.

Kundennummer	Name	Vorname	Anrede	PLZ	Ort	Straße
0001	Dampf	Hansjörg	Herr	00100	Musterort	Hauptstr. 4
0002	Petersen	Griselda	Frau	00123	Beispielstadt	Festplatz 9

Abbildung 1

Die fett gedruckten Überschriften sind die einzelnen Feldbezeichnungen. Alle Daten in einer Spalte sind von derselben Art (zum Beispiel nur Nachnamen oder Postleitzahlen). Die einzelnen Zeilen fassen jeweils die Daten eines Kunden zusammen und werden als Datensätze bezeichnet.

Natürlich kann man mit dieser Kundenliste allein noch nicht viele Auswertungen vornehmen, da sie keinerlei Betragsspalten oder ähnliches enthält. Aber hier ging es zunächst um das Prinzip (und auch solche Tabellen mit Stammdaten haben ihren Nutzen und können etwa mit den Tabellen, die die Umsätze enthalten, verknüpft werden).

Wichtig ist, dass jeder Datensatz eine eindeutige Kennung hat. Dies ist in den meisten Fällen eine fortlaufende Nummer, wie im obigen Beispiel die Kundennummer. Der Nachname alleine wäre nicht ausreichend als Unterscheidungsmerkmal, da Namen wie „Müller“ oder „Maier“ in einer größeren Kundendatei sicher mehrmals vorkommen. Selbst die Kombination aus Nachnamen und Vornamen wäre nicht immer eindeutig, könnte es doch beispielsweise mehrere Hans Müller geben.

Diese eindeutige Kennzeichnung, der sogenannte Primärschlüssel, wird auch verwendet, um zwei Tabellen miteinander zu verbinden.

2.2. Power Pivot

Während PowerQuery für den Import und die Aufbereitung der Daten zuständig ist, erfolgt die eigentliche Analyse mit dem neuen und sehr mächtigen Tool Power Pivot. Power Pivot ist Bestandteil von Excel (allerdings nicht in allen Versionen; beispielsweise ist es nicht in der Home- and Student-Edition enthalten, sondern nur in den größeren Editionen. Vor dem Kauf von Excel sollte man sich auf der Website von Microsoft erkundigen, ob Power Pivot tatsächlich enthalten ist oder gegebenenfalls zu einer höheren Version gegriffen werden muss. Meines Erachtens empfiehlt es sich nicht, zugunsten eines niedrigeren Preises auf Power Pivot zu verzichten, selbst dann nicht, wenn man aktuell noch keine konkreten Anwendungsmöglichkeiten

dafür sieht. Diese ergeben sich, sobald man sich näher mit den Möglichkeiten von Power Pivot befasst.

Zunächst müssen die Daten, die wiederum in einer Datenbankstruktur vorliegen müssen, in Power Pivot importiert werden⁵.

Obwohl auch Power Pivot selbst einige Importmöglichkeiten hat, wird in den meisten Fällen die Reihenfolge diejenige sein, dass zunächst mit PowerQuery der Import und die Aufbereitung der Daten erfolgen und diese dann von PowerQuery an das Datenmodell von Power Pivot übergeben werden.

Bemerkenswert ist, dass die Beschränkung „normaler“ Excel-Arbeitsblätter auf circa eine Million Zeilen⁶ nicht gilt. Die Größe von Tabellen, das heißt die Anzahl der Datensätze, wird nur von den physikalischen Rahmenbedingungen (Speicherkapazität und Rechenleistung des Computers) begrenzt.

Innerhalb von Power Pivot stehen zahlreiche Möglichkeiten der Aufbereitung und Auswertung von Daten zur Verfügung. Beispielsweise kann man – ähnlich wie in Excel selbst – Formeln verwenden. Diese werden mit einer eigenen Formelsprache mit Namen „DAX“ formuliert. Ebenfalls werden in Power Pivot die KPIs zur Kennzahlenanalyse definiert.

Besonders hervorzuheben ist die Möglichkeit, verschiedene Datentabellen zu verknüpfen (anhand ihrer Schlüsselfelder). So kann man übergreifende Auswertungen über mehrere Datentabellen vornehmen und anschaulich in Pivottabellen zusammenfassen.

Beispielsweise könnte die Liste der Umsätze die Personalnummer des jeweiligen Vertreters enthalten. Der Name des Vertreters findet sich in einer anderen Tabelle, der Personaldatenbank. Über die Personalnummer lässt sich eine Verbindung herstellen, die den einzelnen Umsätzen den Vertreternamen zuordnet.

2.2.1. Pivottabellen

Pivottabellen⁷ gibt es schon länger im „normalen“ Excel, Microsoft hat sie aber mit Power Pivot auf ein ganz neues Niveau gehoben. Ein Beispiel (Abbildung 2/nächste Seite) soll die Funktionsweise von Pivottabellen demonstrieren.

Es handelt sich bei diesem einfachen Beispiel um die Raumliste einer Immobilienverwaltung. Man erkennt die Datenbankstruktur, wobei jede Spalte Informationen derselben Art enthält (etwa die Nutzungsart) und jede Zeile als Datensatz einen Raum bezeichnet. Alle Informationen, die einen Raum definieren, müssen in jedem Datensatz enthalten sein, also hier Lage, Nutzung, Geschoss usw.

⁵ Tatsächlich werden die Datentabellen zum neuen „Datenmodell“ von Excel hinzugefügt. Dieses ist sozusagen die im Hintergrund arbeitende Datenbank von Power Pivot. Dabei werden die Daten technisch gesehen in einem sogenannte „Datenwürfel“ oder „Data Cube“ abgelegt, was an dieser Stelle aber nicht weiter von Belang sein soll.

⁶ Das hört sich nach viel an, ist aber schnell erreicht, wenn man beispielsweise eine komplette Fibu über mehrere Jahre einlesen muss.

⁷ Pivot ist Französisch und bedeutet „Dreh- und Angelpunkt“. Diese Bezeichnung verweist auf die Eigenschaft von Pivottabellen, dass die einzelnen Felder beliebig angeordnet werden können, wobei jede Anordnung eine völlig andere Sicht auf die Daten ermöglicht und damit wieder ganz neue Informationen zutage fördert.

Bei den normalen Excel-Pivottabellen müssen tatsächlich alle Informationen in einer einzigen Tabelle stehen, und gleichzeitig muss sich die gesamte Tabelle in einem Excel-Arbeitsblatt befinden (was auch bedeutet, dass die Grenze von einer Million Zeilen nicht überschritten werden darf).

Tabellen im Datenmodell dürfen dagegen beliebig groß sein, und die Daten können außerdem in verschiedenen Tabellen vorliegen, die sich über die Schlüsselfelder miteinander verknüpfen lassen.

Gebäude	Geschoss	Raumnr.	Fläche	Nutzung
Stuttgart	UG	UG_01	40,00 m²	Lager
Stuttgart	UG	UG_02	60,00 m²	Lager
Stuttgart	EG	EG_01	30,00 m²	Produktion
Stuttgart	EG	EG_02	70,00 m²	Büro
Stuttgart	OG1	OG1_01	40,00 m²	Büro
Stuttgart	OG1	OG1_02	30,00 m²	Wohnen
Stuttgart	OG1	OG1_03	30,00 m²	Wohnen
Leipzig	UG	UG_01	120,00 m²	Lager
Leipzig	UG	UG_02	80,00 m²	Lager
Leipzig	EG	EG_01	60,00 m²	Büro
Leipzig	EG	EG_02	60,00 m²	Büro
Leipzig	OG1	OG1_01	40,00 m²	Büro
Leipzig	OG1	OG1_02	40,00 m²	Büro
Leipzig	OG1	OG1_03	40,00 m²	Büro
Leipzig	OG2	OG2_01	60,00 m²	Büro
Leipzig	OG2	OG2_02	60,00 m²	Büro
München	UG	UG_01	90,00 m²	Lager
München	UG	UG_02	80,00 m²	Lager
München	EG	EG_01	50,00 m²	Produktion
München	EG	EG_02	120,00 m²	Produktion
München	OG1	OG1_01	20,00 m²	Büro
München	OG1	OG1_02	60,00 m²	Büro
München	OG1	OG1_03	90,00 m²	Lager
München	OG2	OG2_01	100,00 m²	Wohnen
München	OG2	OG2_02	70,00 m²	Wohnen

Abbildung 2

Diese Raumliste ist zwar relativ klein, dennoch ist sie zu schon zu umfangreich, um daraus auf einen Blick sehen zu können, wie viele Quadratmeter Produktionsfläche im Gebäude München vorhanden sind.

Hier kommt die Pivottabelle ins Spiel. Eine der möglichen Ausgabeform ist in Abbildung 3 unten dargestellt.

Sofort wird ersichtlich, dass in München 170 m² Produktionsfläche vorhanden sind, aber auch, dass insgesamt im Gebäude München 680 m² Fläche zur Verfügung stehen oder in allen Gebäuden zusammen 560 m² Lagerfläche.

Ein Doppelklick auf eines der Wertfelder, das in diesem Fall jeweils die Summe der Flächen aller Räume wiedergibt, für die die Kriterien (Produktion, München) zutreffen, liefert eine Liste der einzelnen Datensätze zutage, aus denen sich diese Zahl zusammensetzt (sogenannte Extraktion oder Drill down). In unserem Beispiel ist das also eine Liste aller Räume, die im Gebäude München liegen und deren Nutzung „Produktion“ ist. Das sieht dann so aus wie in Abbildung 4.

Gebäude	Geschoss	Raumnr.	Fläche	Nutzung
München	OG1	OG1_02	60	Büro
München	OG1	OG1_01	20	Büro

Abbildung 4

Dieser Beitrag konnte nur versuchen, das Interesse am Thema BI zu wecken. Die Möglichkeiten sind fast unbegrenzt und es lohnt, einen genaueren Blick zu wagen. Übrigens ist eine Pivottabelle auch der Dreh- und Angelpunkt einer Betriebsprüfung. Der Prüfer ordnet in der Software IDEA die Daten in einer Pivottabelle an, deren Spaltenbeschriftungen die Jahre sind und die Zeilenbeschriftungen die einzelnen Konten. Ein Doppelklick auf einen Saldo liefert gewissermaßen einen Kontoauszug.

Sum of Fläche	Column Labels				Grand Total
Row Labels	Büro	Lager	Produktion	Wohnen	
Leipzig	360,00 m²	200,00 m²			560,00 m²
München	80,00 m²	260,00 m²	170,00 m²	170,00 m²	680,00 m²
Stuttgart	110,00 m²	100,00 m²	30,00 m²	60,00 m²	300,00 m²
Grand Total	550,00 m²	560,00 m²	200,00 m²	230,00 m²	1.540,00 m²

Abbildung 3

Seminar des Autors

Grundvermögensbewertung nach der ErbSt-Reform

Referenten: Stephan Blum, Dipl.-Finanzwirt (FH) und Wolfgang Weiss, Dipl.-Fw. (FH), StB

25.06.2018, 09:00–16:30 Uhr, München, LSWB-Akademie, Hansastrasse 32

26.06.2018, 09:00–16:30 Uhr, Nürnberg, Hotel Arvena Park, Görlitzer Straße 51