Einführung in die Datenwissenschaft mit Excel

Christian Glahn

Inhaltsverzeichnis

Vc	Vorwort				
Co	ppyright	7			
I.	Ausgangslage und Vorbereitung	8			
1.	Einleitung 1.1. Ziele 1.2. Organisation dieses Buchs	9 9			
2.	Excel Versionen 2.1. Excel-Versionen und Betriebssysteme	10 11 12			
3.	Bedienoberfläche 3.1. Arbeitsblätter 3.2. Arbeitsblattreiter 3.3. Menubalken 3.4. Funktionsleiste 3.5. Statusleiste 3.5.1. Zoom der Tabellenanzeige 3.5.2. Status und Fehlermeldungen 3.5.3. Schnellauswertung eines Zellenbereichs 3.5.4. Anpassen der Statusleiste	14 15 15 16 17 17 17 18 18			
4.	Excel Jargon	19			
II.	Datenquellen	21			
5.	Daten sammeln 5.1. Daten mit Formularen sammeln	22 22 23 24 24			

		5.1.4. Einfachauswahl ($Auswahl$)					24
		5.1.5. Mehrfachauswahl (Auswahl)					24
		5.1.6. Rangfolge					25
		5.1.7. Likert-Skala ($Likert$)					25
		5.1.8. Bewertung					25
		5.1.9. Net Promoter Score					25
		5.1.10. Darstellung der Ergebnisse in Excel					26
	5.2.						26
		5.2.1. Schemadefinition durch Datenüberprüfung					27
		5.2.2. Funktionsweise der Datenüberprüfung					28
		5.2.3. Datenüberprüfung für Zahlen, Zeichenketten oder Wahr	rheit	swe	rte		30
		5.2.4. Datenüberprüfung für ganze Zahlen					30
		5.2.5. Datenüberprüfung für ordinalskalierte Wertebereiche .					32
		5.2.6. Datenüberprüfung für nominalskalierte Wertebereiche					32
		•					
6.		entypen					35
	6.1.	Fundamentale Datentypen					35
		6.1.1. Zahlen					35
		6.1.2. Zeichenketten					35
		6.1.3. Wahrheitswerte					36
		6.1.4. Fehlerwerte					36
		6.1.5. Formeln					37
		6.1.6. Leere Zellen					37
	6.2.	Dekoratoren					37
		6.2.1. Apostroph-Dekoraktor					38
		6.2.2. Gleich-Dekorator					38
	6.3.	1					38
		6.3.1. Bereiche: Vektoren und Matrizen					39
		6.3.2. Tabellen					39
	6.4.	Adressierung von Datenstrukturen					40
		6.4.1. Arbeitsblattadressen					40
		6.4.2. Tabellenadressen					42
,	D-4						46
<i>(</i> .		ten importieren					46
		8					48
	7.2.	T.					51
		7.2.1. Datentyp anpassen					52
		7.2.2. Spalten umbenennen					53
		7.2.3. Spalten verschieben					55
	7.0	7.2.4. Spalte teilen					55
	7.3.	•					56
	7.4.		• •				56
	7.5	Datenverbindung annassen					57

Ш	I. Mathematik der Daten	60
8.	Variablen, Funktionen und Operatoren	61
	8.1. Variablen	61
	8.1.1. Dynamische Bereiche	61
	8.1.2. Benannte Bereiche	62
	8.2. Funktionen	63
	8.3. Operatoren	63
	8.3.1. Substitution über Funktionspfade	63
	8.3.2. Substitution mit LET()	63
	8.4. Funktionsketten	63
	8.4.1. Leere Zellen in einer Funktionskette	63
9.	Zeichenketten	65
10). Aussagenlogik	66
	10.1. Wahrheitswerte in Excel	66
	10.2. Aussagenlogische Operationen	67
	10.3. Vergleiche	69
	10.3.1. Der ∈-Operator mit XVERWEIS	69
	10.3.2. Zeichenketten vergleichen	70
	10.4. Komplexe logische Ausdrücke und Datenstrukturen	70
	10.5. Fälle unterscheiden	72
	10.5.1. WENN	72
	10.5.2. WENNS	73
	10.5.3. Nicht erreichbare Entscheidungen	77
	10.5.4. ERSTERWERT	78
	10.5.5. XVERWEIS zur Fallunterscheidung	79
	10.5.6. Anwendungshilfe für Fallunterscheidungen	80
	10.6. Filtern	80
	10.6.1. Excel Filter und logische Operationen	82
	10.7. Selektieren	82
	10.7. Selektieren	
		84 86
	10.9. Rezepte	
	10.9.1. Fehlerwerte abfangen	86
	10.9.2. Eine Zahl für genau eine Bedingung zurückgeben	86
	10.9.3. Fehlerwerte vergleichen	88
	10.9.4. Filtern und Summen	89
11	. Vektoroperationen	90
12	2. Matrix-Operationen	91
13	3.Kodieren und Gruppieren	92

14. Daten formen	93
IV. Deskriptive Datenanalyse	94
15. Daten beschreiben	95
16. Daten visualisieren	96
17. Pivot-Tabellen	97
Referenzen	98

Vorwort

Work in Progress

Copyright

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz (CC-BY-NC-SA). Details zur Nutzungsbedingungen und dem Copyright finden sich unter createivecommons.org.

2023, Christian Glahn, Zurich, Switzerland



Teil I. Ausgangslage und Vorbereitung

1. Einleitung



1.1. **Z**iele

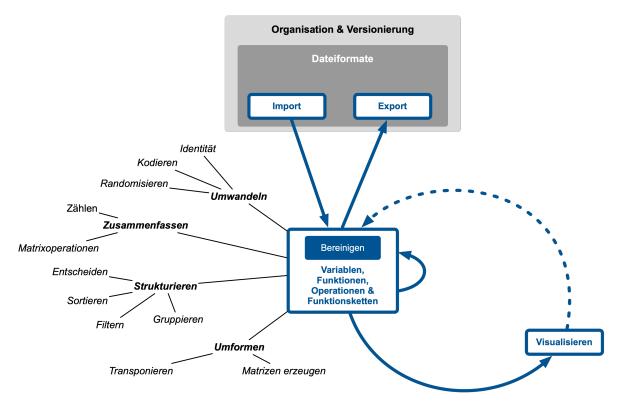


Abbildung 1.1.: Ziele und Themen

1.2. Organisation dieses Buchs

2. Excel Versionen

Excel ist die Tabellenkalkulation der Microsoft Office Umgebung. Die erste Version wurde 1985 veröffentlicht und wird seitdem ständig weiterentwickelt. Aktuell existieren mehrere Versionen von Excel nebeneinander.

- Excel365
- Excel für Desktop-Computer ohne Microsoft365
- Excel als Web-Anwendung
- Excel für Mobilgeräte

Die verschiedenen Versionen unterscheiden sich in der Funktionalität.

Die aktuelle Version ist **Excel365** und muss auf einem Desktop-Computer oder Laptop installiert werden. Diese Version ist Teil des Microsoft365 Abonnements. Diese Version wird kontinuierlich weiterentwickelt und erhält regelmässig neue Funktionen, welche die Arbeit erleichtern.

Teil des Microsoft365 Abonnements sind die Web- und die Mobilversionen von Excel. Beide Versionen sind in der Funktion leicht eingeschränkte Versionen von Excel365. Die Mobilversion ist für die Arbeit auf einem Smartphone oder Tablet optimiert. Die Web-Version benötigt einen Web-Browser und kann auf jedem Gerät mit Internetzugang verwendet werden. Der Unterschied zwischen Excel365 und den wird immer kleiner und die meisten Konzepte in diese Buch lassen sich auch in Excel im Web umsetzen.

Excel's Desktop Version ohne Abonnement wurde auf dem Stand von Juni 2019 eingefrohren. Diese Version wird nur noch mit Sicherheitsupdates versorgt. Neue Funktionen sind in dieser Version nicht verfügbar. Diese Version ist noch recht weit verbreitet und lässt sich nur schwer von der aktuellen Version unterscheiden, denn tatsächlich handelt es sich bei Excel365 und der Desktop-Version ohne Abonement um die selbe Anwendung. Fehlt das Abonnement, dann sind die neuen Funktionen nicht verfügbar und werden nicht angezeigt oder ausgeführt. Falls diese Version aktiv ist und eine Arbeitsmappe, die mit einer Version mit Abonement erstellt wurde, öffnet, dann erfolgt eine Fehlermeldung und die entsprechenden Funktionen werden blockiert.

Dieses Buch setzt die aktuelle Version von Excel365 voraus. In den folgenden Kapiteln ist also immer Excel365 gemeint, wenn schlicht von Excel geschrieben wird. Die meisten Beispiele und Konzepte lassen sich in der Desktop-Version *ohne Abonnement* nicht umsetzen. Viele Beispiele lassen sich mit der Web- oder Mobilversion nachvollziehen, diesen Versionen fehlt allerdings

der Importer für Daten aus anderen Dateien, der in Kapitel 7 behandelt wird und für viele Beispiele Voraussetzung ist.

Ist die richtige Excel Version installiert

Mit einem einfachen Test lässt sich überprüfen, ob eine geeignete Excel Version installiert ist:

- 1. Eine beliebige Zelle mit der Maus ausgewählt.
- 2. Es werden die drei Zeichen '=HS eingegeben.

Nun sollte der Funktionsvorlschag für die Funktion HSTAPELN unterhalb der Eingabe erscheinen. Wird kein Funktionsvorschlag angezeigt, dann ist eine zu alte Version von Excel installiert.

Die Eingabe wird mit ESC abgebrochen.

Erscheint kein Funktionsvorschlag, dann muss ein zweiter Test erfolgen.

- 1. Eine beliebige Zelle mit der Maus ausgewählt.
- 2. Es werden die Zeichen =seq eingegeben.

Nun sollte der Funktionsvorlschag für die Funktion SEQUENZ unterhalb der Eingabe erscheinen. In diesem Fall muss nur ein kleines Update von Microsoft365 installiert werden.

Erscheint kein Funktionsvorschlag, dann ist eine veraltete Version von Excel installiert. Diese muss zuerst vollständig deinstalliert werden, bevor die neue Version von Microsoft365 installiert werden kann. Beim vollständigen Deinstallieren leider müssen auch alle Einstellungen aller Microsoft365 Anwendungen entfernt werden.

2.1. Excel-Versionen und Betriebssysteme

Excel365 unterscheidet sich leicht für MacOS und Windows. Im Rahmen dieses Buchs stellt das keine Einschränkung dar, denn die Unterschiede sind gering.

Die auffälligsten Unterschiede betreffen den Importer und die bedingte Formatierung.

- Der Importer ist unter Windows etwas flexibeler gestaltet als unter MacOS. Dieser Unterschied betrifft den Inhalt dieses Buchs nicht.
- Die bedingte Formatierung wird unter MacOS traditionell mit einen völlig andereren Dialog konfiguriert wird als unter Windows.

A MacOS vs. Windows

In den (wenigen) Fällen, in denen sich die Excel-Versionen auf den beiden Betriebssystemen unterscheiden, wird der Unterschied durch eine MacOS vs. Windows-Box - wie diese hier - gesondert hervorgehoben. Das gleiche gilt für die Eigenheiten, die nur auf einem der beiden Betriebssyeme vorkommen.

⚠ MacOS

Soll unter MacOS Excel365 eine alte Version ohne Abonement ersetzen, dann muss die alte Version zuerst deinstalliert werden. Sind zusätzlich andere Office-Programme von Microsoft ohne Abonement installiert, dann müssen diese ebenfalls deinstalliert werden. Beim Deinstallieren werden die Konfigurationsdateien nicht gelöscht. Diese müssen manuell gelöscht werden. Erst jetzt kann die neue Version von Microsoft365 installiert werden.

2.2. Excel und Sprachen

Excel365 ist in vielen Sprachen verfügbar. Die Sprache wird von Excel vom Betriebssystem übernommen. Dieses Buch bezieht sich auf die deutschsprachige Excel Version mit Schweizer Regionseinstellungen. Alle Beispiele lassen sich in allen anderen Sprachen nachvollziehen, wenn die Werte und Funktionsnamen entsprechend angepasst werden.



Warnung

Die Funktionsnamen unterscheiden sich stark zwischen den einzelnen Sprachen. Oft ist es unmöglich, die Funktionsnamen durch eine Übersetzung des Funktionsnamen aus einer anderen Sprache abzuleiten.

Beispielsweise heisst der englische Funktionsname OFFSET() (deutsch: Versatz) auf Deutsch Bereich. Verschieben() (engl. move range).



A MacOS vs. Windows

Unter MacOS kann die Sprache von Excel365 ausschliesslich über die Systemeinstellungen geändert werden. Die Spracheinstellung wird vom Betriebssystem übernommen. Dadurch verändert sich die Sprache aller Programme, die auf dem Computer installiert sind. Wird die Sprache von MacOS geändert, dann muss der Computer neu gestartet werden, damit die Änderung wirksam wird. Dieser Schritt verändert jedoch nicht nur die Sprache von Menus sondern auch die Namen vieler Verzeichnisse. Deshalb wird dieser Schritt nicht empfohlen.

Unter Windows kann die Sprache nur für Excel
365 geändert werden. Dazu muss in Excel unter Datei > Optionen > Sprache geändert werden. Die gewünschte Sprache muss unter Office-Anwendungssprache als bevorzugt markiert werden. Die Änderung wird erst wirksam, nachdem Excel beendet und neu gestartet wurde.

Die Spracheinstellung haben keinen Einfluss auf die Funktion von Excel. Die Spracheinstellung bestimmt nur die Sprache der Benutzeroberfläche. Zur Benutzeroberfläche gehören die Menüs, die Dialoge, die Hilfetexte und die **Funktionsnamen**. Wird eine Arbeitsmappe in einer Excel mit einer anderen Spracheinstellung geöffnet, dann werden die Funktionsnamen automatisch übersetzt. Es ist jedoch unmöglich die Funktionsnamen in der ursprünglichen Sprache der Arbeitsmappe einzugeben.

3. Bedienoberfläche

3.1. Arbeitsblätter

Der Tabellenbereich ist das zentrale Element von Excels Arbeitsoberfläche. Dabei handelt es sich um den Datenbereich, der sich deutlich durch das beschriftete Gitterraster im Zentrum des Excel-Fensters abhebt und den Grossteil der Arbeitsoberfläche einnimmt. Über dieses Raster wird primär mit Daten und Funktionen interagiert. Der Tabellenbereich zeigt einen Ausschnitt des aktuellen Arbeitsblatts.

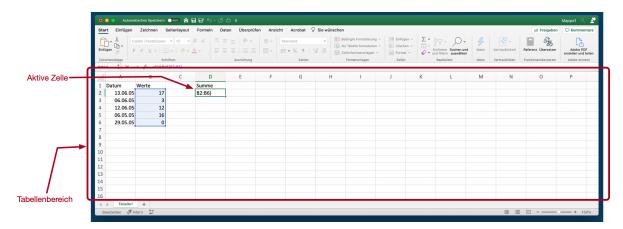


Abbildung 3.1.: Das Excel Arbeitsfenster

Definition 3.1. Ein Excel Arbeitsblatt ist ein Gitterraster, das aus Zellen besteht. Jede Zelle kann einen Wert enthalten.

Definition 3.2. Eine **Zelle** ist ein Feld im Raster eines Arbeitsblatt.

Excel zeigt durch eine Hervorhebung an, welche Zelle aktuell markiert ist. Diese **Markierung** kann mit der Maus oder mit den Pfeiltasten bewegt werden. Wenn nur eine Zelle markiert ist, dann entspricht die Markierung der **aktiven Zelle**. Eine Markierung kann auch mehrere Zellen umfassen. In diesem Fall wird die *aktive Zelle* durch eine helle Hervorhebung gekennzeichnet, während der Rest der Markierung ausgegraut ist. Alle Eingaben mit der Tastatur bezeihen sich immer auf die aktive Zelle.

Oben und Links vom Tabellenbereich findet sich die Beschriftung für die Spalten und Zeilen. Spalten werden mit Buchstaben benannt, Zeilen werden mit Zahlen benannt. Daraus ergibt sich ein Koordinatensystem, mit dem jede Zelle eindeutig adressiert wird.

Definition 3.3. Die Koordinaten einer Zelle sind die Zellenadresse.

Im Abschnitt 6.4 werden die verschiendenen Arten der Adressierung von Zellen und Zellbereichen detailliert beschrieben.

3.2. Arbeitsblattreiter

Direkt unter dem Tabellenbereich zeigt Excel Register mit den Namen der Arbeitsblätter. Diese Register zeigen alle Arbeitsblätter einer Excel-Arbeitsmappe. Wenn eine neue, leere Arbeitsmappe erstellt wird, dann enthält diese nur ein Arbeitsblatt mit dem Namen Tabelle 1. Mit dem Plus-Symbol neben dem Arbeitsblattregistern können zusätzliche Arbeitsblätter in der Arbeitsmappe erzeugt werden.

3.3. Menubalken



Abbildung 3.2.: Menubalken

Am oberen Fensterrand von Excel befindet sich der **Menubalken**, der im Excel Jargon "Ribbon" heisst. Der Menubalken ist in verschiedene Kategorien unterteilt und sortiert so die verschiedenen interaktiven **Kommandos** von Excel.

Definition 3.4. Ein **Kommando** ist eine *interaktive Anweisung* an Excel, eine bestimmte Aktion auszuführen.

Kommandos werden immer interaktiv ausgelöst und lassen sich meist nicht durch Excel-Funktionen ersetzen.

Für die tägliche Arbeit sind die folgenden Kategorien des Menubalkens am wichtigsten:

- Start
- Einfügen
- Formeln
- Daten
- Tabelle

3.4. Funktionsleiste

Die Funktionsleiste unterstützt die Arbeit mit der aktiven Zelle. Die Funktionsleiste wird verwendet, wenn mit speziellen Werten oder mit Komplexen Formeln gearbetiet wird.

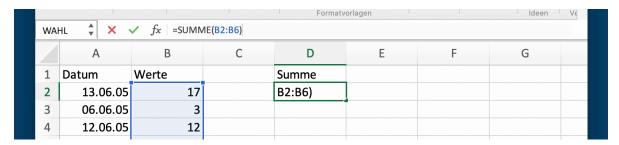


Abbildung 3.3.: Die Excel Funktionsleiste

Die Funktionsleiste hat 6 Bedienfelder.

Ganz links ist das **Adressfeld**. Das *Adressfeld* zeigt uns immer die Adresse der *aktiven Zelle* an. Wenn wir dieses Feld anklicken, dann können wir eine andere Adresse eintragen. In diesem Fall springt die Markierung für die aktive Zelle an die entsprechende Adresse.

Daneben befinden sich zwei unscheinbare Dreiecke. Dieses Bedienfeld verbirgt eine Liste der benannten Bereiche in der aktuellen Arbeitsmappe. Bei einer leeren Arbeitsmappe ist diese Liste leer.

Die nächsten vier Bedienfelder gehören zusammen.

Ganz rechts wird der Inhalt der aktuellen Zelle angezeigt. Bei einer leeren Tabelle ist dieser Bereich selbstverständlich leer. Wenn dieser Bereich mit der Maus ausgewählt wird, dann erscheint ein senkrechter, blinkender Strick, der signalisiert, dass Werte oder Formeln eingegeben werden können. Wenn nun etwas eingegeben wird, dann erscheint diese Eingabe auch in der aktiven Zelle.

Ist die Eingabe abgeschlossen, dann kann diese mit dem grünen Haken oder mit der Eingabetaste bestätigt werden.

Ist bei der Eingabe ein Fehler passiert, dann lässt sich die Eingabe mit einem Mausklick auf das rote X oder mit der ESC-Taste abbrechen. Die Eingabe wird dann auf den ursprünglichen Zustand zurückgesetzt.

Wird die aktive Zelle nicht bearbeitet, dann sind der Hacken und das X ausgegraut und können nicht verwendet werden.

Direkt neben dem Eingabefeld mit dem Inhalt der aktiven Zelle ist das Bedienfeld für den Formelgenerator. Der Formelgenerator ist ein Hilfsdialog, der die Suche nach geeigneten Formeln unterstützt. Der Formelgenerator unterscheidet sich von Menubalken Formeln durch eine kurze Erklärung über die Arbeitsweise einer ausgewählten Funktion.

3.5. Statusleiste

Am unteren Rand von Excels Anwendungsfenster findet sich die Statusleiste. Die Statusleiste hat drei zentrale Funktionen:

- Zoom der Tabellenanzeige,
- Status von Excel und Fehlermeldungen und
- Schnellauswertung eines Zellbereichs.

3.5.1. Zoom der Tabellenanzeige

Der aktuelle Zoom der Arbeitsmappe und die Zoom-Einstellung finden sich ganz rechts in der Statusleiste. Dabei wird der aktuelle Zoom als ein Prozentwert dargestellt. Über den "Schieberegler" links daneben lässt sich der Zoom anpassen.

Wenn der Zoom-Wert angeklickt wird, dann erscheint ein kleiner Dialog, über denden Zoom-Wert als Prozentzahl direkt eingegeben kann. Das ist praktisch, wenn für eine Präsentation oder zum Arbeiten ein bestimmter Zoom-Wert bevorzugt wird.



Achtung

Die Zoom-Einstellung muss für jeddes Arbeitsblatt separat eingestellt werden und wird beim nächsten Speichern der Arbeitsmappe mitgesichert.

3.5.2. Status und Fehlermeldungen

Auf der linken Seite der Statusleiste ist der Status Bereich. Normalerweise wird an dieser Stelle den Wert Bereit angezeigt. Wenn eine Zelle bearbeitet wird, dann erscheint der Status Eingeben oder Zeigen.

Wenn einen Fehler bei der Eingabe einer Formel auftritt, dann wird neben dem Status eine Fehlermeldung eingeblendet. Diese unscheinbare Fehlermeldung wird immer dann wichtig, wenn eine Formel oder ein Algorithmus nicht zu rechnen scheint. In solchen Fällen hilft ein Blick auf den linken Rand der Statusleiste bei der Fehlersuche. Bei Fehlern zeit die Statusleiste einen Hinweis auf den Fehler sowie die Adresse der Zelle, an der Excel den Fehler vermutet.

Tritt ein Fehler auf, dann bricht Excel die Berechnung der gesamten Arbeitsmappe sofort ab und ignoriert alle Formeln, die nach der fehlerhaften Zelle berechnet werden müssten, selbst wenn die Formeln unabhängig voneinander wären. Damit Excel wieder arbeitet, müssen zuerst alle Fehler behoben werden.

3.5.3. Schnellauswertung eines Zellenbereichs

Im mittleren Bereich der Statusleiste ist normalerweise leer. Hier versteckt sich eine sehr nützliche Funktion zur schnellen Zusammenfassung der Werte in einem Bereich: Werden mehrere Zellen markiert, dann zeigt Excel den Mittelwert, die Summe und die Anzahl der ausgewählten Zellen an.

3.5.4. Anpassen der Statusleiste

Wenn Sie mit der rechten Maustaste (Win) bzw. mit Control-Klick (Mac) auf die Statusleiste klicken, dann erhalten Sie eine Auswahlliste der Funktionen, die in der Statusleiste angezeigt werden sollen.

So lassen sich zusätzliche Funktionen in der Statusleiste aktivieren oder störende Elemente deaktivieren.



⚠ Warnung

Diese Einstellungen beziehen sich ausschliesslich auf die Excel-Installation und werden nicht zusammen mit der Arbeitsmappe gespeichert!

4. Excel Jargon

Die Dokumentation von Excel verwendet eine Reihe von Begriffen, die von den Begriffen der Datenwissenschaften und der Informatik abweichen. Dieses Kapitel enthält ein Glossar dieser Begriffe.

Arbeitsmappe: Bezeichnet eine Excel-Datei.

Arbeitsblatt: Als Arbeitsblatt wird in der Regel eine Tabelle in einer Excel-Arbeitsmappe bezeichnet.

Adresse: Koordinaten einer Zelle oder eines Bereichs auf einem Arbeitsblatt.

Aktive Zelle: Die Zelle, die für das Bearbeiten ausgewählt ist. Ist nur eine Zelle markiert, dann entspricht die aktive Zelle der Markierung. Ist ein Bereich markiert, dann ist die aktive Zelle, die weiss hervorgehobene Zelle.

Bereich: Eine Auswahl von Zellen auf dem gleichen Arbeitsblatt.

dynamischer Bereich oder dynamisches Feld oder dynamisches Array: Ein Bereich mit Werten, der durch eine Formel erzeugt wird.

Bezug: Adresse eines Bereichs, um den Wert bzw. die Werte an der Adresse zu erhalten.

Externe Daten: Tabellen, die aus einer anderen Datei oder einer Datenbank eingebunden wurden.

Formel: Aufruf von Funktionen in einer Zelle.

Funktion: Bestandteil einer Formel. Funktionen werden über Parameter gesteuert.

Kommando: Ein Kommando ist eine Interkaktion, mit der Excel eine Aktion auslöst. Aktionen ähneln Funktionen, können aber nicht als Formel ausgedrückt werden.

Parameter: Wert, der an eine Funktion übergeben wird. Ein Parameter kann direkt als Wert, indirekt als Adresse oder als Ergebnis einer anderen Funktion übergeben werden.

Pivot-Tabelle: Eine Kreuztabelle, mit der Daten aus mehreren Spalten eines Arbeitsblatts zusammengefasst werden können. Eine Pivot-Tabelle kann Daten von mehreren Arbeitsblättern und Tabellen zusammenfassen.

Ribbon: Menüleiste von Excel.

Tabelle: Eine Tabelle ist eine zwei-dimensionale Datenstruktur für strukturierte Daten. Eine Tabelle besteht aus Spalten und Zeilen.

Zelle: Ein Feld in einer Tabelle. Eine Zelle besteht immer aus einem Werteanteil und einem Formelanteil. Normalerweise wird der Werteanteil angezeigt.

Teil II. Datenquellen

5. Daten sammeln

5.1. Daten mit Formularen sammeln

Die direkte Eingabe von Daten ist nur für kleine Datensätze mit wenigen Merkmalen geeignet. Bei grösseren Datensätzen ist die direkte Eingabe oft zu aufwändig und fehleranfällig. Für die strukturierte Dateneingabe von grösseren Datensätzen werden Formulare verwendet.

Bei der Dateneingabe über Formulare werden die Daten mithilfe von *Fragen* mit vorgegebenen Antworttypen gesammelt. Der Antworttyp bestimmt, welche Werte für die Frage zulässig sind.

i Hinweis

In vielen Formularsystemen wird der *Antworttyp* als *Fragetyp* bezeichnet, obwohl die Frage immer in Text-Bildform präsentiert wird und der *Typ* durch die unterschiedlichen Antwortmöglichkeiten bestimmt wird.

Durch das Festlegen des Antworttyps wird gleichzeigt der Wertebereich des gemessenen Merkmals definiert. Dadurch ergibt sich beim Erstellen eines Formulars das Datenschema automatisch.

Excel hat kein eingenes Formularsystem. Stattdessen kann Microsoft Forms verwendet werden. Microsoft Forms ist ein Online-Formulardienst, der in Microsoft365 integriert ist. Forms sammelt die Eingaben mit einem Formular und speichert die Daten in einer Excel Arbeitsmappe. Diese Formulare können über einen Web-Link geteilt werden und im Web-Browser ausgefüllt werden. Dadurch kann das Datensammeln auf verschiedene Personen verteilt werden.

i Hinweis

Ein Formular kann in Microsoft Forms ein **Quiz** oder ein **Formular** sein. Ein Quiz ist ein Formular, bei dem die Antworten mit einer Punktzahl bewertet werden und ein Feedback erhalten können. Ein Formular fokussiert sich auf die reine Dateneingabe.

Microsoft Forms unterstützt die folgenden Antworttypen:

• Offene Antworttypen

- Offene Texteingabe
- Dateumseingabe
- Datei hochladen
- Geschlossene Antworttypen
 - Einfachauswahl
 - Mehrfachauswahl
 - Rangfolge
 - Likert-Skala
 - Bewertung
 - Net Promoter Score

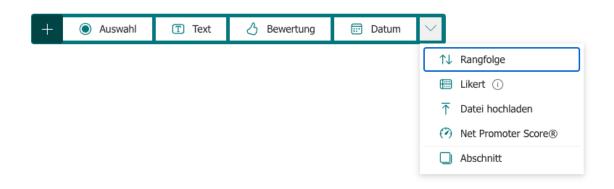


Abbildung 5.1.: MS Forms Fragetypmenu

Jede dieser Varianten hat die folgenden gemeinsamen Optionen:

- Eine einleitende Frage oder Anweisung.
- Ein optionaler Untertitel oder eine Beschreibung. (Untertitel)
- Die Möglichkeit eine mathematische Formel zur Validierung des Wertebereichs anzugeben (*Mathematik*).
- Das erzwingen einer Antwort (Erforderlich).
- Eine Verzweigung zu einer anderen Frage, wenn eine bestimmte Antwort gegeben wird (*Verzweigung*).

Die Unterschiede ergeben sich in den Antwortmöglichkeiten.

5.1.1. Offene Texteingabe (Text)

Die offene Texteingabe erlaubt einzeilige und mehrzeilige (lange) offene Antworten möglich sind. Damit können un- oder semistrukturierte Daten erfasst werden. Diese Antwortmöglichkeit lässt sich auf Zahlenwerte und Zahlenintervalle eingeschränken.

5.1.2. Datumseingabe (Datum)

Diese Eingabe ist eine Variante der offenen Texteingabe zur Eingabe eines Datums. Diese Eingabe wird durch die Option ergänzt, das Datum über eine Kalenderdarstellung auszuwählen.

Das eingegebene Datum wird als Zahl gespeichert und als Datum dargestellt.

5.1.3. Datei hochlanden

Die zweite Variante für offene Antworten ist das Hochladen von Dateien. Mit diesem Antworttyp lassen sich Bilder, Videos oder Tonaufnahmen leicht erfassen.

5.1.4. Einfachauswahl (*Auswahl*)

Die Einfachauswahl ist ein Single-Choice-Antworttyp. Es sind nur die vorgegebenen Antwortmöglichkeiten möglich, von denen nur eine ausgewählt werden kann. Die Antwortmöglichkeiten können als Text oder als Bild angegeben werden. Die Antwortmöglichkeiten werden immer in einer Liste angezeigt. Dabei kann zwischen einem Auswahlmenu als Dropdown-Liste und frei anwählbaren Antwortmöglichkeiten (Optionen) gewählt werden.

Die Reihenfolge der Antwortmöglichkeiten kann durch die Option Optionen in zufälliger Reihenfolge für jedes Formular zufällig angeordnet werden. Diese Option ist sinnvoll, wenn die Reihenfolge der Antwortmöglichkeiten die Antwort beeinflussen könnte. Soll ein Formular mehrfach durch die gleichen Personen ausgefüllt werden, dann sollte diese Option nicht aktiviert werden.

Das Ergebnis der Einfachauswahl ist eine Zeichenkette der ausgewählten Option.

5.1.5. Mehrfachauswahl (Auswahl)

Die Mehrfachauswahl erlaubt die Auswahl mehrerer Antwortmöglichkeiten, indem die Option Mehrere Antworten aktiviert wird. Dieser Antworttyp ist ansonsten identisch mit der Einfachauswahl.

Das Ergebnis der Mehrfachauswahl ist eine Zeichenkette mit den ausgewählten Optionen. Wurden mehrere Antworten ausgewählt, dann werden die Optionen durch ein Semikolon getrennt.

5.1.6. Rangfolge

Die Rangfolge erlaubt es, dass die angebotenen Antwortmöglichkeiten in eine Reihenfolge gebracht werden können. Bei diesem Antworttyp müssen die Werte sortiert werden und die Sortierung wird als Antwort gespeichert.

Bei der Darstellung werden die Antwortmöglichkeiten in einer Liste angezeigt und für jedes Formular neu sortiert.

Das Ergebnis der Rangfolge ist eine Zeichenkette mit der Reihung der Optionen. Die Optionen werden durch ein Semikolon getrennt. Im Ergebnis lässt sich die Rangfolge nicht von einer Mehrfachauswahl unterscheiden.

5.1.7. Likert-Skala (Likert)

Der Antworttyp Likert setzt eine Fragebatterie mit Likert-Skalen um. Darüber lässt sich ein semantisches Differential erheben.

Die Anworten dieser Fragebatterie werden wie mehrere Einfachauswahlen als Zeichenketten gespeichert. Deshalb muss für diesen Antworttyp immer eine Kodierungstabelle für die richtige Reihenfolge der Werte dokumentiert werden.

5.1.8. Bewertung

Die Bewertung ist eine Variante des Antworttyps *Likert-Skala*. Mit der Bewertung wird über eine festgelegte Skala der gewünschte Wert abgefragt. Optional können für die beiden Extremwerte Beschriftungen angegeben werden.

Das Ergebnis wird als Zahl gespeichert. Die Zahl entspricht dem Wert der ausgewählten Option beginnend bei 1.

5.1.9. Net Promoter Score

Der Typ Net Promoter Score ist eine spezielle elfstufige Liker-Skala, die zwischen zwei frei wählbaren Extremwerten festgelegt ist. Die offiziell als *Net Promoter Score* (Reichheld, 2003) bezeichnete Eingabeform ist eine verzerrte Likert-Skala mit dem Mittelpunkt bei 7.5.

Der Net Promoter Score wird immer als Ganzzahl im Intervall $0 \leq s \leq 10$ zurückgegeben.

⚠ Warnung

Der Net Promoter Score hat im Management-Umfeld eine grosse Verbreitung gefunden. Die Verwendung dieser Skala ist aber umstritten (Keiningham et al., 2007). Speziell die proklamierte Aussagekraft lässt sich nicht belegen (Fisher & Kordupleski, 2019; Grisaffe, 2007). Von der verbreiteten isolierten Verwendung des Net Promoter Score nach der Methode von Reichheld (2003) wird deshalb abgeraten.

Wird diese Eingabe als Likert-Skala verwendet, dann besteht ausser bei der Darstellung kein Unterschied zur Einfachauswahl oder der Bewertung.

5.1.10. Darstellung der Ergebnisse in Excel

Die Ergebnisse der Formulare werden in einer Excel-Arbeitsmappe als Tabelle gespeichert. Die Antworten sind in der Spalte ID sequenziell durchnummeriert. Zusätzlich enthält die Tabelle zwei sog. Zeitstempel mit Datum und Uhrzeit des Beginns (Startzeit) und des Endes (Fertigstellungszeit) der Datenerfassung eines Formulars. Optional können die Namen und E-Mail-Adressen der ausfüllenden Person miterfasst werden.

Nach diesen Basisinformationen folgen die einzelnen Antworten für die Formularelemente. Die Werte in der gleichen Zeile entsprechen dabei den Werten aus dem gleichen Formular. Die Spaltenüberschriften sind Standardmässig der Fragetext.

5.2. Daten direkt eingeben

Bei der direkten Eingabe von Daten muss zuerst die Ergebnistabelle manuell erstellt werden, um damit die Daten zu sammeln.

Damit die Daten einheitlich erfasst werden, benötigt die Tabelle ein Schema. In Excel wird diese Schema über die Tabellenüberschriften definiert und kann bei einer späteren Verarbeitung weiter angepasst werden (s. Kapitel 7).

Ein einfaches Datenschema umfasst nur die Spaltenüberschriften, die erfasst werden sollen. Dazu müssen die Tabellenüberschriften definiert werden. Die Tabellenüberschriften werden in der ersten Zeile der Tabelle eingetragen. Überschriften für weitere Merkmale können zu einem späteren Zeitpunkt noch ergänzt werden. Anschliessend werden die Überschriften markiert und es wird für diesen Bereich eine **Tabelle** eingefügt. Hierbei muss die Option **Tabelle** hat Überschriften aktiviert werden (s. Abbildung 5.2).

Daten werden der Tabelle hinzugefügt, indem in der Zeile unterhalb der Tabelle neue Werte eingetragen werden. Die Tabellenstruktur mit dem Schema wird automatisch erweitert.

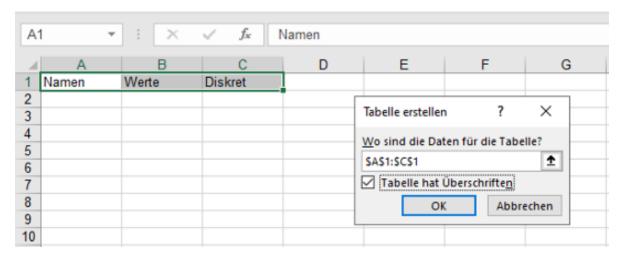


Abbildung 5.2.: Tabelle erstellen

5.2.1. Schemadefinition durch Datenüberprüfung

Die einfache Schemadefinition stellt nicht sicher, dass die Werte einheitlich abgelegt werden. Das kann später die Datenerfassung und die Datenverarbeitung behindern. Wurde die Datentabelle mit einem Überprüfungsschema erstellt, dann können die Werte in die Tabelle eingetragen und bei der Eingabe validiert werden.

Die Schemadefinition durch Datenüberprüfung erlaubt die Wertebereiche der gemessenen Merkmale vorab festzulegen. Dadurch lassen sich potentielle Fehler bereits bei der Dateneingabe vermeiden. Hierzu ist es notwendig, die **Wertebereiche** und die zugehörige **Datenklasse** einer Spalte zu bestimmen.



Das Erstellen eines Schemas zur Dateneingabe ist in Excel zwar aufwändig, dennoch sollte dieser Schritt nicht übersprungen werden, weil die Datenüberprüfung die Datenqualität verbessert und die Datenverarbeitung erheblich vereinfacht.

In Excel müssen Vektoren durch die Option *Datenüberprüfung* definiert werden. Die Option *Datenüberprüfung* findet sich im Ribbon unter *Daten* (s. Abbildung 5.3) im Abschnitt *Datentools* (s. Abbildung 5.4).



Abbildung 5.3.: Ribbon Datenüberprüfung

Die Schemadefinition erfolgt in drei zusätzlichen Schritten:

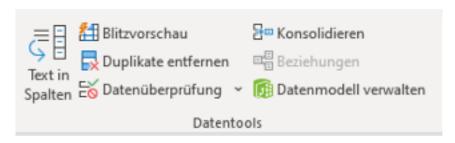


Abbildung 5.4.: Datentools Details

- 2. Es wird eine *Platzhalterzeile* eingefügt, indem die Zelle unterhalb der ersten Überschrift ausgewählt wird. Die Platzhalterzeile wird benötigt, um die Datenüberprüfung zu definieren.
- 3. Für jeden Vektor wird mit der Option *Datenüberprüfung* eine Datenüberprüfung für den gewünschten Wertebereich definiert.
- 4. Die Platzhalterzeile wird gelöscht.

5.2.2. Funktionsweise der Datenüberprüfung

Die Datenüberprüfung ist eine Funktion, die auf einen Bereich angewendet wird, wobei der Bereich eine einzelne Zelle, einer Zeile oder eine Spalte sein kann. Für das Schema einer Tabelle wird die Datenüberprüfung für eine Spalte eingerichtet.

Die Datenüberprüfung wird Zellenweise konfiguriert, wobei die erste Zelle eines Bereichs die Referenzzelle ist. Die Datenüberprüfung wird auf die Referenzzelle angewendet und diese Konfiguration wird auf alle Zellen des Bereichs übertragen und automatisch so angepasst, dass die aktuelle Zelle überprüft wird. Dabei sind ein paar Besonderheiten zu beachten.

- Die Datenüberprüfung muss mit konstanten Adressen konfiguriert werden. Es muss also die doppelte Dollar-Adressierung (z.B. \$A\$2) verwendet werden.
- Wird nicht die Referenzzelle als Adresse verwendet, dann erfolgt die Datenüberprüfung mithilfe der Werte der angegebenen Adresse. In den folgenden Zellen wird dann der relative Versatz zur Referenzzelle verwendet.
- Konstante Werte müssen in der Datenüberprüfung explizit angegeben und nicht über Bezüge verwiesen werden.

Excel bringt einige vordefinierte Überprüfungen mit, die über die Dropdown-Liste ausgewählt werden können. Die meisten Überprüfungen sind für Zahlenintervalle definiert. Die Standardeinstellung Jeden Wert ist gleichbedeutend mit keiner Überprüfung, weil alle Werte als Eingabe akzeptiert werden. Generelle Überprüfungen von Datentypen sind nicht vorgegeben und müssen über die Option Benutzerdefiniert definiert werden.

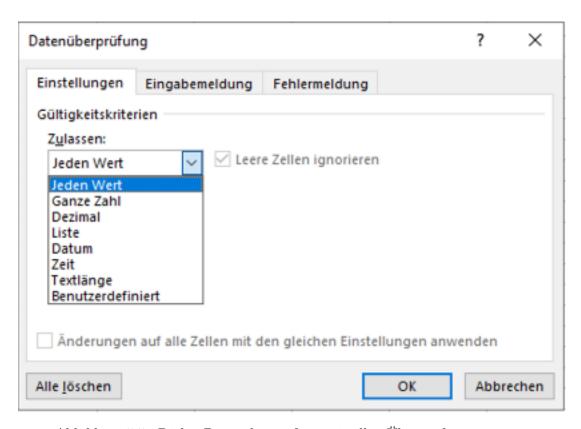


Abbildung 5.5.: Dialog Datenüberprüfung mit allen Überprüfungsoptionen

i Hinweis

Wird die Datenüberprüfung auf eine Zelle in einer Tabelle angewendet und in der jeweiligen Spalte existieren noch keine anderen Werte, dann wird die Datenüberprüfung für die gesamte Spalte übernommen. Stehen in der Tabelle bereits Werte, dann müssen zuerst alle Werte in der Spalte ausgewählt werden, bevor die Datenüberprüfung eingerichtet wird. Neue Werte übernehmen die Datenüberprüfung der vorangehenden Zeile.

5.2.3. Datenüberprüfung für Zahlen, Zeichenketten oder Wahrheitswerte

Eine generelle Überprüfung für die fundamentalen Datentypen ist in Excel nicht vorgesehen. Solche Überprüfungen müssen über die Option Benutzerdefiniert festgelegt werden.

Im Eingabefeld Formel kann eine beliebig komplexe Formel stehen. Die Formel muss einen Wahrheitswert zurückgeben. Wird der Wahrheitswert WAHR zurückgegeben, dann wird der Wert akzeptiert. Wird der Wahrheitswert FALSCH zurückgegeben, dann wird der Wert nicht akzeptiert.

Um die fundamentalen Datentypen zu überprüfen, können die Informationsfunktionen ISTZAHL(), ISTTEXT() und ISTLOG() verwendet werden (s. Tabelle 5.1). Diese Funktionen geben WAHR zurück, wenn der Wert dem angegebenen Datentyp entspricht. Ansonsten geben sie FALSCH zurück.

Tabelle 5.1.: Überprüfungsausdrücke für die fundamentalen Datentypen

Datentyp	Überprüfungsausdruck
Zahlen	=ISTZAHL(\$A\$2)
Zeichenketten	=ISTTEXT(\$A\$2)
Wahrheitswerte	=ISTLOG(\$A\$2)

5.2.4. Datenüberprüfung für ganze Zahlen

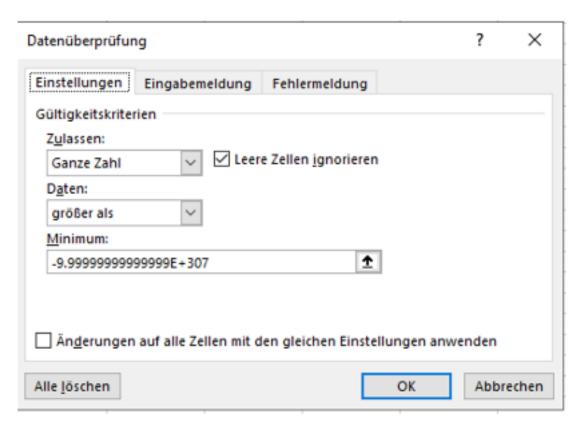


Abbildung 5.6.: Dialog Datenüberprüfung für alle ganzen Zaheln

5.2.5. Datenüberprüfung für ordinalskalierte Wertebereiche

Für ordinalskalierte Wertebereiche sollten die Werte am Besten als Ganzzahlen erfasst werden. Damit zu einem späteren Zeitpunkt diese Zahlen den eigentlichen Werten zugeordnet werden können, wird eine Zuordnungstabelle benötigt. Diese Zuordnungstabelle wird in einer separaten Tabelle erstellt. Diese Tabelle besteht aus zwei Spalten: Dem Zahlenwert und dem zugeordneten Wert. Die Ordnung der Werte ergibt sich aus der Ordnung der zugeordneten Zahlen.

Wichtig

Häufig muss bei ordinalskalierten Wertebereichen die Option einer nicht zutreffenden Antwort berücksichtigt werden. In diesem Fall muss eine zusätzlicher Zahlenwert ausserhalb der regulären Ordnung definiert werden. Dieser Wert wird dann der nicht zutreffenden Antwort (z.B. NA oder nicht zutreffend) zugeordnet.

Damit der Wertebereich korrekt eingeschränkt wird, muss die Option Liste verwendet werden. Die Quelle entspricht der Spalte mit den zugeordneten Zahlenwerten in der Zuordnungstabelle. Die Option Zellendropdown kann deaktiviert werden.

Achtung

Die Datenüberprüfung kann nicht mit Tabellenadressen umgehen. Deshalb müssen die gültigen Werte als absolute Arbeitsblattadressen angegeben werden (s. Abbildung 5.7).

5.2.6. Datenüberprüfung für nominalskalierte Wertebereiche

Die Überprüfung nominalskalierter Wertebereiche wird in Excel durch die Option Liste realisiert. Die Option Liste erlaubt die Auswahl von Werten aus einer Liste in der gleichen Arbeitsmappe. Diese Liste ist einfach ein Bereich mit Werten, eine Spalte in einer anderen Tabelle oder eine benannte Liste (s. Kapitel 8). Die Liste muss in einer Spalte definiert werden, weil die Datenüberprüfung nur auf Spalten angewendet werden kann. Deshalb ist es nicht möglich, eine zeilenorientierte Liste für die Datenüberprüfung anzuwenden.

Für nominalskalierte Wertebereiche werden alle zulässigen Werte in einer Spalte gestgelegt. Anschliessend wird im Dialog Datenüberprüfung unter Quelle der Bereich mit den zulässigen Werten angegeben.



Tipp

ordinalskalierten Wertebereichen, sollten nominalskalierte Wertebereiche nicht als Zahlen kodiert werden, sondern als Zeichenketten erfasst werden.

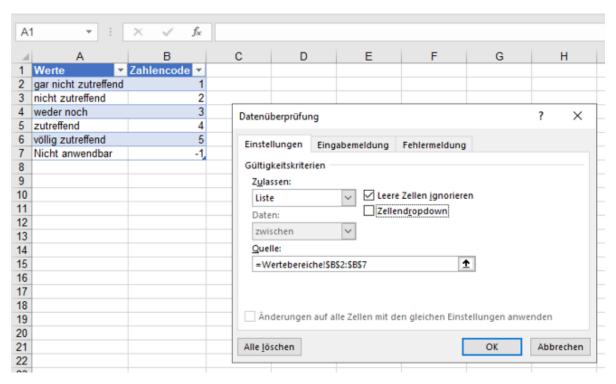


Abbildung 5.7.: Datenüberprüfung mit Zuordnungstabelle für ordinalskalierte Wertebereiche

Bei der Dateneingabe wird anschliessend ein Auswahlliste mit den zulässigen Werten, so dass die Werte nicht mehr direkt eingetippt werden müssen. Diese Auswahlliste kann unterbunden werden, wenn die Option Zellendropdown deaktiviert wird.

6. Datentypen

6.1. Fundamentale Datentypen

Excel kennt sechs fundamentale Datentypen.

- Zahlen
- Zeichenketten
- Wahrheitswerte
- Fehlerwerte
- Formeln
- leere Zelle

6.1.1. Zahlen

Zahlen werden in Excel immer als Gleitkommazahlen behandelt. Excel kennt von keine ganzen Zahlen. Wenn eine Zahl in eine ganze Zahl umgewandelt wird (z.B. durch die Funktion GANZZAHL), wird lediglich der Nachkommaanteil der Zahl auf 0 gesetzt.

Manche Excel Funktionen arbeiten nur mit ganzen Zahlen. In diesen Fällen wird der Nachkommaanteil der Zahl automatisch abgeschnitten.

Durch die Verwendung von Gleitkommazahlen können in Excel Zahlen mit einer Genauigkeit von 15 signifikanten Stellen dargestellt werden. Werden zwei Zahlen addiert, und das Ergebnis mehr als 15 signifikante Stellen hätte, werden die alle Stellen ab der 15. signifikanten Stelle abgeschnitten. Excel versucht diese Fehler möglichst zu vermeiden.

Der Datentyp Zahlen wird in Excel mit der Funktion ISTZAHL() geprüft. Die Funktion ISTZAHL() liefert WAHR, wenn der Wert ein Zahl ist, und sonst FALSCH.

6.1.2. Zeichenketten

Zeichenketten heissen in Excel **Text**. Zeichenketten werden von Excel automatisch gewählt, wenn kein anderer Datentyp für eine Eingabe erkannt wurde. Um die automatische Erkennung zu verhindern, muss der Apostroph-Dekorator (s. Abschnitt 6.2.1) verwendet werden.

Im Formelmodus müssen Zeichenketten in doppelte Anführungszeichen eingeschlossen werden. Im Wertemodus müssen Zeichenketten meist nicht besonders markiert werden.

Der Datentyp Zeichenkette wird in Excel mit der Funktion ISTTEXT() geprüft. Die Funktion ISTTEXT() liefert WAHR, wenn der Wert eine Zeichenkette ist, und sonst FALSCH.

Besondere Zeichenketten sind Adressen bzw. im Excel Jargon **Bezüge**. Die Funktion ISTBEZUG() prüft, ob eine Zeichenkette eine gültige Excel Adresse ist. Die Funktion ISTBEZUG() liefert WAHR, wenn die Zeichenkette eine Adresse ist, und sonst FALSCH.

6.1.3. Wahrheitswerte

Wahrheitswerte heissen in Excel WAHR und FALSCH. Wahrheitswerte heissen im Excel Jargon logische Werte. Während der Werteeingabe werden Wahrheitswerte unabhängig von der Gross- und Kleinschreibung automatisch erkannt.

Im Formelmodus dürfen Wahrheitswerte *nicht* in Anführungszeichen eingeschlossen werden, weil sonst die Symbole als Zeichenkette behandelt werden.

Der Datentyp Wahrheitswerte wird in Excel mit der Funktion ISTLOG() geprüft. Die Funktion ISTZLOG() liefert WAHR, wenn der Wert ein Wahrheitswert ist, und sonst FALSCH.

6.1.4. Fehlerwerte

Fehlerwerte ist in Excel ein besonderer Datentyp, um Fehler zu signalisieren. Fehler beginnen immer mit einem Gatter (#), das von einem sog. Fehlerbezeichner gefolgt wird. In der Regel werden Fehlerwerte automatisch erzeugt, es ist aber möglich, Fehlerwerte auch manuell einzugeben. Bei der manuellen Eingabe von Fehlerwerten werden nur die gültigen Fehlerbezeichner akzeptiert. Andere Symbolfolgen werden als Zeichenketten interpretiert.

Excels gültige Fehlerwerte sind #NV, #NULL!, #WERT!, #BEZUG!, #DIV/O!, #ZAHL!, #NAME?, #KALK!, #ÜBERLAUF!, #DATEN_ABRUFEN! und #ZAHL!.

Der Wert #NV wird in Excel verwendet, um einen ungültigen Wert zu verweisen. Dieser Wert ist nicht gleichbedeutend mit fehlenden Werten.

Der Datentyp Fehlerwert wird in Excel mit der Funktion ISTFEHLER() geprüft. Die Funktion ISTFEHLER() liefert WAHR, wenn der Wert ein Fehlerwert ist, und sonst FALSCH. Um einen bestimmten Fehlerwert zu prüfen, kann die Funktion FEHLER.TYP() mit dem entsprechenden Fehlerbezeichner als Argument verwendet werden. Diese Funktion gibt einen eindeutigen Zahlenwert für den Fehler zurück. Ist der Wert kein Fehlerwert, dann wird der Fehler #NV zurückgegeben.

6.1.5. Formeln

Formeln sind ein eigener Datentyp, die mit dem Gleich-Dekorator (=) beginnen. Eine Formel besteht immer aus einem Wert, einem Funktionsaufruf oder einer Kombination von Werten und Funktionsaufrufen, die durch Operatoren verknüpft wurden.

Der Datentyp Formel wird in Excel mit der Funktion ISTFORMEL() geprüft. Die Funktion ISTFORMEL() liefert WAHR, wenn der Wert eine Formel ist, und sonst FALSCH.

Formeln sind in Excel ein besonderer Datentyp, denn eine Zelle mit einer Formel hat immer zwei Datentypen. Der erste Datentyp ist der Datentyp Formel, der zweite Datentyp ist der Datentyp des Formelergebnisses.

6.1.6. Leere Zellen

Lehre Zellen sind Zellen, die keinen Wert enthalten. Leere Zellen zeigen fehlende Werte in Excel an.

Die Leere Zelle ist für Excel ein eingener Datentyp und wird mit der Funktion ISTLEER() geprüft. Die Funktion ISTLEER() liefert WAHR, wenn der Wert eine leere Zelle ist, und sonst FALSCH.



Warnung

Eine leere Zelle und eine leere Zeichenkette lassen sich mit dem Auge nicht unterscheiden. Die Funktion ISTLEER() liefert für eine leere Zeichenkette FALSCH zurück, weil die Zelle einen Wert enthält.

Leere Zellen als Ergebnis einer Formel werden von Excel in die Zahl 0 umgewandelt. Im Kapitel Abschnitt 8.2 wird gezeigt, wie sich leere Zellen von Zellen mit dem Wert 0 unterscheiden lassen. Im Abschnitt 8.4.1 wird auf leere Zellen als Funktionsergebnisse ausführlich eingegangen.

6.2. Dekoratoren

Definition 6.1. Ein **Dekorator** ist Sprachelement einer Programmiersprache, mit dem die normale Interpretation von Symbolen verändert werden kann.

Excel hat zwei Dekoratoren, als erstes Symbol einer Zelle stehen müssen. D.h. es dürfen auch keine Leerzeichen vor einem Dekorator stehen. Die beiden Dekoratoren sind:

- Der Apostroph-Dekorator (')
- Der Gleich-Dekorator (=)

6.2.1. Apostroph-Dekoraktor

Der Apostroph-Dekorator (') erzwingt, dass die nachfolgenden Symbole als Zeichenkette interpretiert werden müssen.

Mithilfe des Apostroph-Dekorators wird die automatische Typerkennung für die laufende Eingabe deaktiviert. Auf diese Weise lassen sich Zeichenketten eingeben, die nur aus Ziffern bestehen oder Zeichenketten, die Datentumswerten ähneln.

Der Apostroph-Dekorator muss zwingend verwendet werden, wenn auf eine Zeichenkette eine der folgenden Bedingungen zutrifft.

- Die Zeichenkette beginnt mit einem Gleichheitszeichen (=), einem Pluszeichen (+), einem Minuszeichen (-) oder einem Prozentzeichen (%).
- Die Zeichenkette besteht nur aus Ziffern, dem Dezimaltrenner (.) oder dem Tausendertrenner (') enthält
- Die Zeichenkette ähnelt der wissenschaftliche Notation für Zahlen mit und ohne Vorzeichen für den Exponenten. Z.B. 5.E3.
- Die Zeichenkette ähnelt einm Datum.
- Die Zeichenkette entspricht einen Fehlerwert, unabhängig von der Schreibweise. Z.B. #nv.
- Die Zeichenkette entspricht einem Wahrheitswert (WAHR oder FALSCH) unabhängig von der Schreibweise.

Das Apostroph kann entfallen, wenn die aktive Zelle vorher bereits als Datentyp Text formatiert wurde. In diesem Fall verlieren alle Symbole ihre spezielle Bedeutung. In solche Zellen können keine Formeln eingegeben werden.

6.2.2. Gleich-Dekorator

Der Gleich-Dekorator (=) zeigt an, dass die nachfolgenden Symbole als Formel interpretiert werden müssen.

Bei der Eingabe des Gleich-Dekorators wechselt Excel für die laufende Eingabe in den Formelmodus. Mit beeenden der Eingabe mit der Eingabetaste wird der Formelmodus wieder verlassen und das Ergebnis der eingegebenen Formel wird in der Zelle angezeigt.

6.3. Komplexe Datenstrukturen

Excel kennt zwei komplexe Datenstrukturen:

- Bereiche
- Tabellen

Excels komplexe Datenstrukturen müssen einen fundamentalen Datentyp haben und können keine komplexen Datenstrukturen schachteln.

6.3.1. Bereiche: Vektoren und Matrizen

Excels Grundstruktur ist das Rechteck, dass durch die markierten Zellen entsteht. Dieses Rechteck heisst im Excel Jargon ein **Bereich**.

Obwohl Excel vektorähnliche Bereiche kennt, ist es sinnvolle sich Excels Bereiche immer als Matrizen vorzustellen.

- Ein Bereich mit nur einer Zelle ist eine 1x1- oder 0-dimensionale Matrix.
- Ein Bereich mit nur einer Zeile oder einer Spalte ist für Excel eine 1xn- oder 1-dimensionale Matrix. Diese speziellen Matrizen werden in der Regel als *Vektoren* bezeichnet.
- Ein Bereich mit mehreren Zeilen und Spalten ist eine nxm- oder 2-dimensionale Matrix.

Funktionen können Bereiche als Ergebnis haben. In diesem Fall werden die restlichen Werte zeilen und spaltenweise unterhalb bzw. rechts von der entsprechenden Formel ausgegeben.

6.3.2. Tabellen

Tabellen sind benannte Bereiche, die Vektoren enthalten. Tabellen bestehen aus Spalten, in denen die Werte als Vektoren stehen. Excel Tabellen haben immer Überschriften.

Excel kennt zwei Arten von Tabellen:

- Wertetabellen
- Pivot-Tabellen.

Definition 6.2. Wertetabellen sind Listen, die Listen mit gleicher Länge schachteln.

Wertetabellen entsprechen ungefähr einem *Datenrahmen* (engl. *data-frame*) in anderen Programmiersprachen. Der Unterschied zu einem Datenrahmen ist, dass Excels Wertetabellen *keine Vektoren* erzwingen. Deshalb können in der gleichen Spalte einer Wertetabelle verschiedene Datentypen gemischt werden.

Definition 6.3. Pivot-Tabellen sind ein Werkzeug zum *interaktiven Zusammenfassen* von Daten.

Pivot-Tabellen erleichtern einfache Datenanalysen. Pivot-Tabellen entsprechen einer tabellarischen Darstellung der Daten. Ihre Funktion ist auf wenige analytische Funktionen beschränkt und die dargestellten Werte lassen sich nur umständlich weiterverarbeiten.

Pivot-Tabellen sind Arbeitsmappenelemente aber **keine Datenstrukturen**. Deshalb werden Pivot-Tabellen in diesem Buch ähnlich einer Visualisierung als Darstellungsform behandelt.

6.4. Adressierung von Datenstrukturen

Definition 6.4. Ein **Bezug** ist die Adresse eines Bereichs.

In Excel gibt es zwei Arten von Bezügen:

- Arbeitsblattadressen
- Tabellenadressen

6.4.1. Arbeitsblattadressen

Definition 6.5. Eine **Arbeitsblattadresse** enthält die Koordinaten eines Bereichs auf einem Arbeitsblatt.

Arbeitsblattadressen besteht aus drei Teilen:

- Arbeitsblattname
- Bereichsbeginn
- Bereichsende

Der Bereichsbeginn verweist immer auf die linke obere Zelle des Bereichs. Der Bereichsende verweist immer auf die rechte untere Zelle des Bereichs. Eine Zelle wird immer durch den Spaltenindex (Buchstabe) und den Zeilenindex (Zahl) identifiziert. Werden die Koordinaten des Bereichsbeginns und des Bereichsendes bei der Eingabe vertauscht, dann wird der Bereich automatisch korrigiert.

Beispiel 6.1 (Arbeitsblattadresse).

Tabelle1!A1:C3

Beispiel 6.1 verweist auf den Bereich mit drei Spalten und drei Zeilen auf dem Arbeitsblatt Tabelle1 beginnend mit der Zelle A1 und endend mit der Zelle C3.

Oft werden Arbeitsblattadressen nicht vollständig sondern gekürzt angegeben. Es gibt zwei Möglichkeiten, um Arbeitsblattadressen zu kürzen:

- Werden Bereiche auf dem gleichen Arbeitsblatt adressiert, dann kann der Arbeitsblattname weggelassen werden.
- Wird ein Bereich mit nur einer Zelle adressiert, dann wird das Bereichsende weggelassen.

Weil Arbeitsblattadressen von vielen interaktiven Excelkommandos verwendet werden, gibt es zwei Arten von Arbeitsblattadressen:

- Relative Adressen
- Absolute Adressen

Die Art der Adresse legt fest, wie ein interaktives Kommando mit einer Adresse umgehen soll. Die populärste interaktive Funktion ist das Autoauffüllen. Dabei wird eine Zelle mit einer Formel interaktiv auf einen Bereich von Zellen übertragen.

🔔 Warnung

Das Autoauffüllen ist eine einfache und beliebte Methode, um Formeln in Excel auf verschiedene Werte wiederholt anzuwenden. Bis 2019 war das Autoauffüllen die einzige Möglichkeit für die Datentransformation.

Die relative und absolute Adressierung ist eine wichtige Voraussetzung für das Autoauffüllen. Leider ist das Autoauffüllen auch die Ursache für viele Fehler beim Umgang mit Excel.



Praxis

In Excel365 kann das Autoauffüllen durch vektorisierte Funktionen fast vollständig ersetzt werden. Dadurch lassen sich viele Excel-typische Fehler vermeiden. Dadurch ist die Unterscheidung zwischen der relativen und absoluten Adressierung nicht mehr so wichtig.

6.4.1.1. Relative Adressen

Definition 6.6. Eine relative Adresse ist eine Adresse eines Bereichs, die veränderlich ist.

Relative Adressen werden in Excel von interaktiven Funktionen, wie dem Autoauffüllen verwendet, um die Adressen automatisch anzupassen. Eine relative Adresse wird relativ zur aktuellen Zelle angepasst.

Ein Beispiel für eine relative Adresse ist A1. Diese Adresse bezeichnet die Zelle, die sich in der ersten Zeile und der ersten Spalte auf dem aktuellen Arbeitsblatt befindet. Wird die adressierte Zelle interaktiv nach unten aufgefüllt, dann wird die Adresse automatisch zu A2, A3, usw. angepasst. Wird die adressierte Zelle nach rechts aufgefüllt, dann wird die Adresse automatisch zu B1, C1, usw. angepasst.

6.4.1.2. Absolute Adressen

Definition 6.7. Eine **absolute Adresse** ist eine Adresse eines Bereichs, die ganz oder teilweise *unveränderlich* ist.

Der unveränderliche Teil einer Arbeitsblattadresse wird mit einem Dollarzeichen (\$) markiert. Dieser Teil der Adresse wird bei der Anpassung der Adresse nicht verändert. So lassen sich Adressen angeben, die durch interaktive Kommandos nicht verändert werden.

Ein Beispiel für eine absolute Adresse ist \$A\$1. Diese Adresse bezeichnet die Zelle, die sich in der ersten Zeile und der ersten Spalte auf dem aktuellen Arbeitsblatt befindet. Wird die adressierte Zelle interaktiv horizontal oder vertikal aufgefüllt, dann wird die Adresse nicht angepasst.

Auf diese Weise lassen sich konstante Werte in Formeln einbauen.

6.4.2. Tabellenadressen

Spalten und einzelne Werte können über die Tabellenadressierung abgefragt werden (Microsoft Support, 2023b). Das Ergebnis einer solchen Adressierung ist immer ein *dynamisches Feld* bzw. ein *dynamischer Bereich* (s. Abschnitt 8.1.1).

i Hinweis

Jede Tabellenadresse kann auch als Arbeitsblattadresse dargestellt werden. Umgekehrt ist dies nicht möglich.

Eine Tabellenadresse besteht aus zwei Teilen:

- Dem Tabellennamen
- Dem Spaltennamen

Beispiel 6.2 (Tabellenadresse).

Tabelle1[Spalte1]

Das Beispiel 6.2 verweist auf die Spalte Spalte1 der Tabelle Tabelle1.

Wichtig

Die Namen von Tabellen sind *unabhängig* den Arbeitsblattnamen. In der gleichen Arbeitsmappe darf jeweils nur eine Tabelle und nur ein Arbeitsblatt mit dem gleichen Namen existieren. Es ist aber möglich, dass eine Tabelle und ein Arbeitsblatt den gleichen

Namen haben. Das kann zu Verwirrungen führen, denn die Adresse Tabelle1 [Spalte1] und die Adresse Tabelle1!A:A verweisen nicht zwingend auf die gleiche Spalte, denn eine Tabelle muss nicht auf einem Arbeitsblatt mit dem gleichen Namen stehen!

Tabellenadressen können auch gekürzt werden:

- Werden Spalten in der gleichen Tabelle angesprochen, dann kann der Tabellenname weggelassen werden.
- Soll die gesamte Tabelle angesprochen werden, dann kann der Spaltenname weggelassen werden.

6.4.2.1. Tabellenbereiche

Um mehrere Spalten einer Tabelle anzusprechen kann der Bereichsoperator (:) wie bei der Arbeitsblattadressierung verwendet werden. Zusätzlich müssen die Spalten in ein weiteres Paar eckiger Klammern eingeschlossen werden.

Beispiel 6.3 (Tabellenbereich).

Tabelle1[[Spalte1]:[Spalte3]]

Beispiel 6.3 verweist auf alle Spalten zwischen Spalte1 und Spalte3 der Tabelle Tabelle1.



Warnung

Werden zu einem späteren Zeitpunkt neue Spalten zwischen den adressierten Spalten hinzugefügt, dann werden die neuen Spalten automatisch in die Adressierung einbezogen.

Wichtig

Es ist nicht möglich mehrere Spalten einer Tabellen gleichzeitig gezielt zu adressieren, wenn diese nicht unmittelbar nebeneinander stehen. Diese Adressierung ist unmöglich, weil Excel nur rechteckige Bereiche adressieren kann und solche Spalten keinen rechteckigen Bereich bilden.

Die Adressierung aus Beispiel 6.3 kann nicht angepasst werden, so dass nur Spalte1 und Spalte3 adressiert werden, ausser die Spalten Spalte1 und Spalte3 stehen direkt nebeneinander.

6.4.2.2. Diese Zeile-Operator

Einzelne Zeilen einer Tabelle können mit dem Diese Zeile-Operator adressiert werden. Der Diese Zeile-Operator wird mit dem Wert [#Diese Zeile] oder mit @ angegeben. Dieser Operator reduziert die Tabelle auf eine einzelne Zeile und wählt anschliessend die gewünschte Spalte aus. Diese Zeile bedeutet für Excel die aktuelle Zeile des Arbeitsblatts.

Beispiel 6.4 (Diese Zeile-Operator für eine Tabellenzelle).

= Tabelle[@Spalte]

oder

= @Tabelle[Spalte]

Der Operator ist speziell für die Verwendung in Formeln in Tabellenzellen gedacht. Obwohl der Operator in allen Formeln eingesetzt werden kann, sollte der Einsatz auf Formeln beschränkt werden, die sich in einer Tabelle befinden.



Warnung

Der Diese Zeile-Operator kann nicht auf Tabellen angewendet werden, um eine ganze Zeile zu adressieren. Es darf immer nur eine einzelne Zelle adressiert werden. Die Adressierung @Tabelle ist also ungültig.

⚠ Warnung

Der Diese Zeile-Operator ist immer relativ zum Arbeitsblatt und nicht relativ zur aktuellen Tabelle. Es muss nur der Zeilenindex gleich sein, die Zielzelle kann sich auf einem anderen Arbeitsblatt befinden.

Wird der Diese Zeile-Operator in Tabellen unterschiedlicher Länge verwendet, dann können zwei Fälle eintreten:

- 1. Ist die adressierte Tabelle kürzer als die adressierende Tablle, dann wird für die Zeilen, die keine Entsprechung in der adressierten Tabelle haben, der Fehler #WERT! zurückgegeben.
- 2. Ist die adressierte Tabelle länger als die adressierende Table, dann werden alle überzähligen Zeilen ignoriert.

? Praxis

Der Diese Zeile-Operator kann nur auf eine einzelne Zelle in der gleichen Zeile auf einem Arbeitsblatts angewendet werden. Wegen dieser Einschränkung des Diese Zeile-Operators sollten alle Tabellen so gestaltet werden, dass sie in der ersten Zeile des Arbeitsblatts beginnen. Dadurch kann der Diese Zeile-Operator in allen Tabellen verwendet werden.

6.4.2.3. Überschriften adressieren

Eine Excel Tabelle hat immer Spaltenüberschriften. Diese Überschriften können ebenfalls über die Tabellenadressierung adressiert werden. Dazu wird als Spaltenname der Wert #Kopfzeilen verwendet.

Beispiel 6.5 (Kopfzeilen einer Tabelle adressieren).

Tabelle1[#Kopfzeilen]

Um gezielt Kofzeilen zu adressieren, kann die Adressierung aus Beispiel 6.5 mit der Adressierung aus Beispiel 6.3 kombiniert werden (Beispiel 6.6).

Beispiel 6.6 (Kopfzeilen einer Tabelle adressieren).

Tabelle1[[#Kopfzeilen]; [Spalte1]: [Spalte3]]

6.4.2.4. Absolute Adressierung in Tabellen

Ein Tabellenbereich ist immer absolut adressiert. Wird aber nur eine einzelne Spalte einer Tabelle adressiert, dann wird die Adresse relativ addressiert. Damit beim Autoauffüllen diese Adresse nicht verändert wird, muss die Adresse als Bereich angegeben werden.

Beispiel 6.7 (Absolute Tabellenadressierung von einer Spalte).

Tabelle1[[Spalte1]:[Spalte1]]

Wird eine Formel mit dieser Adresse interaktiv aufgefüllt, dann wird die Adresse nicht verändert.

7. Daten importieren

Viele Excel Arbeitsmappen kombinieren Berechnungen und Daten. Excel kann allerdings auch Daten aus anderen Quellen importieren. Quellen können andere Dateien, Datenbanken oder Web-APIs sein und in verschiedenen Formaten vorliegen. Dafür stellt Excel Parser für verschiedene Dateiformate bereit, damit die Daten importiert werden können.

Daten werden korrekt mit dem Kommando **Daten abrufen (Power Query)** importiert. Das Kommando ist in der Gruppe **Daten** im Abschnitt **Daten abrufen und transformieren** zu finden. Das Kommando *Daten abrufen* startet die sog. **Power Query** Umgebung. In dieser Umgebung können Daten aus verschiedenen Quellen importiert und vor der Bereitstellung korrigiert werden.

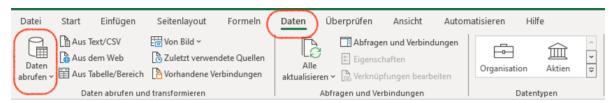
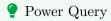


Abbildung 7.1.: Menubalken Power Query - Kommando Daten abrufen



Power Query ist eine Umgebung für den Datenimport, die von verschiedenen Microsoft Produkten verwendet wird. Die Umgebung ist in Excel, Power BI, Power Apps und Power Automate verfügbar. Power Query stellt eine einheitliche Schnittstelle für den Datenimport bereit und basiert auf einer *Import-Beschreibungssprache*. Mit diese Sprache lassen sich Daten für die Arbeit vorbereiten, so dass viele Datenbereinigungsschritte in Excel Arbeitsmappen entfallen können.

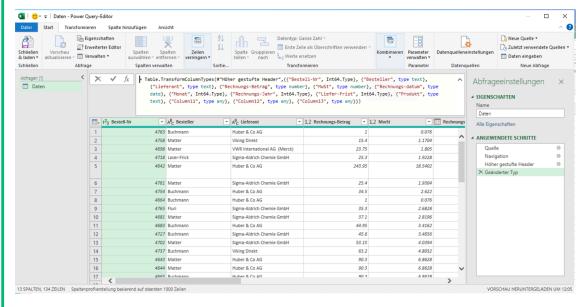


Abbildung 7.2.: Power Query Editor Fenster

Zu den zentralen Funktionen von Power Query gehören:

- Quellenmanagement
- Überschriftenerkennung
- Schemaerkennung und -transformation
- Entfernen von Duplikaten und ungültigen Werten
- Vektorisierung von Daten
- Kombinieren von Daten aus verschiedenen Quellen



Viele tabellarische Dateiformate lassen sich direkt mit Excel öffnen. Das sollte nur mit Excel Arbeitsmappen erfolgen. Bei anderen Dateiformaten kann das direkte öffnen zu Datenverlusten oder Datenfehlern führen. Diese lassen sich in Excel nur umständlich korrigieren.

Ausserdem lassen sich direkt geöffnete Dateien nicht mehr erweitern, was die Datenerhebung erschwert.

Das Kernprinzip von *Power Query* ist das Verbinden einer Datenquelle mit einer Arbeitsmappen. Entsprechend heissen importierte Daten im Excel Jargon **Datenverbindungen**. Die Datenquelle wird dabei nicht in die Arbeitsmappe übernommen, sondern nur eine Verbindung zu der Datenquelle hergestellt. Dadurch kann eine Datenquelle in mehreren Arbeitsmappen verwendet werden und sich ändern, ohne dass die Arbeitsmappe angepasst

werden muss.

i Merke

Das Ergebnis eines Imports ist immer eine Tabelle.

7.1. Datenverbindungen herstellen

Eine Datenverbindung wird über das Kommando Daten abrufen (Power Query) ausgelöst. Das Kommando ist in der Gruppe **Daten** im Abschnitt **Daten abrufen und transformieren** zu finden. Das Kommando bereitet den Import durch Power Query vor, indem zuerst das Dateiformat und anschliessend die zugehörige Datenquellen ausgewählt wird. Nach diesen Schritten kann Excel die Daten Laden oder die Daten transformieren (s. Abbildung 7.3).

Praxis

Die Daten sollten nur in Ausnahmefällen unkontrolliert gelanden werden. Die Daten sollten vor dem Importieren in Power Query überprüft und transformiert werden. Dadurch wird sichergestellt, dass die Daten korrekt importiert werden.

Wird die Option Daten transformieren ausgewählt, öffnet sich der Power Query Editor. In diesem Editor können die Daten überprüft und transformiert werden. Das Transformieren der Daten umfasst vier Arten der Transformation:

- 1. Datentypanpassungen
- 2. Spaltenauswahl
- 3. Datentrennung
- 4. Daten kombinieren

Für jede Transformation wird ein Schritt in Power Query angelegt. Die Schritte werden in der Reihenfolge der Ausführung angezeigt. Die Reihenfolge der Schritte kann verändert werden, indem die Schritte mit der Maus verschoben werden. In der Mitte des Power Query Fensters wird eine Datenvorschau angezeigt, die die Auswirkungen der Transformationen zeigt.

Standardmässig werden drei bzw. bei Excel Arbeitsmappen vier Schritte in Power Query automatisch konfiguriert:

1. Quelle

- Bei Excel Arbeitsmappen folgt dem Schritt Quelle der Schritt Navigation zur Auswahl des Arbeitsblattes.
- 2. Höherstufen
- 3. Geänderter Typ (Windows) bzw. geänderter Spaltentyp (MacOS)

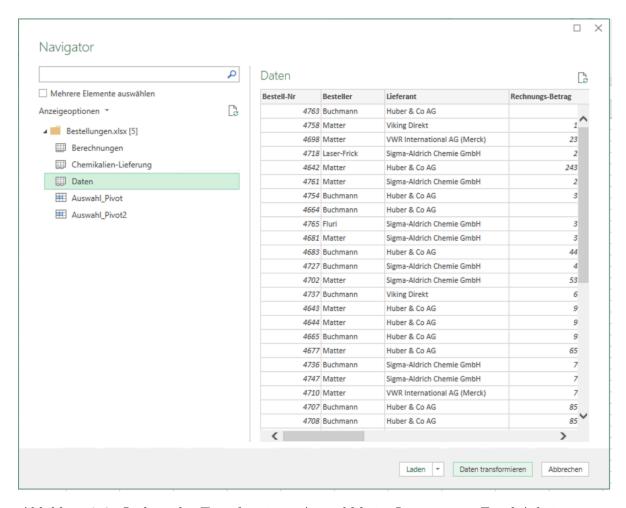


Abbildung 7.3.: Laden oder Transformieren-Auswahl beim Import einer Excel Arbeitsmappe

Der Schritt Geänderter Typ enthält die Schemadefinition. Power Query versucht das Schema automatisch zu erkennen. Das funktioniert besonders bei Zahlenwerten nicht immer zuverlässig. Deshalb muss das Schema immer kontrolliert und gegebenefalls angepasst werden.

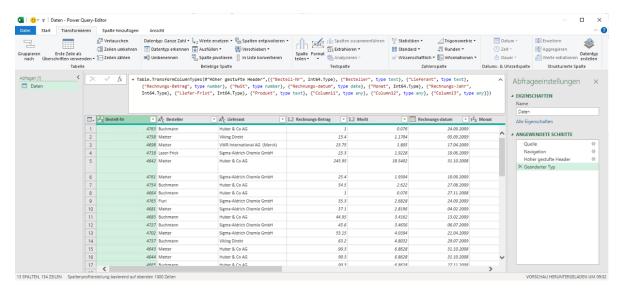


Abbildung 7.4.: Schemaanpassung mit Power Query

Nachdem das Schema kontrolliert wurde, können die Daten nach Excel importiert werden. Dazu wird das Kommando Schließen & laden in Power Query Menukategorie Start ausgeführt.

Nach einem Datenimport liegen die Daten in der Arbeitsmappe als Tabelle vor. Diese Daten sind eine Kopie der Daten in der Datenquelle. Dadurch kann die Arbeitsmappe unabhängig von der Datenquelle verwendet und geteilt werden.



Achtung

betrachtet jeden Datenimport alsSicherheitsproblem. Datenverbindungen beim Öffnen einer Arbeitsmappe standardmässig deaktiviert. Beim Offnen einer Arbeitsmappe mit einer Datenverbindung wird eine Warnung angezeigt. Um mit den importierten Daten arbeiten zu können, muss die Datenverbindung aktiviert werden.

Wird die Datenverbindung nicht aktiviert, werden Funktionen die auf die Daten zugreifen nicht ausgeführt, sondern nur die Ergebnisse der letzten Ausführung angezeigt.

7.2. Datenschema anpassen

Die einfachste Möglichkeit zur Anpassung des Datenschemas ist die Verwendung der Menukategorie Transformation. Dieses Menu zeigt alle verfügbaren Transformationen an. Die Transformationen sind in sechs Kategorien gegliedert:



Abbildung 7.5.: Power Query Menu Transformation

- 1. Tabelle
- 2. Beliebige Spalte
- 3. Textspalte
- 4. Zahlenspalte
- 5. Datums- & Uhrzeitspalte
- 6. Strukturierte Spalte

Für das Datenschema sind die folgenden Transformationen wichtig:

- Datentyp
- Umbenennen
- Verschieben
- Spalte teilen

i Hinweis

Die meisten Transformationen gehören konzeptionell zu späteren Kapiteln.

- Gruppieren nach und Werte ersetzen werden im Kapitel 13 behandelt.
- Vertauschen wird im Kapitel 12 behandelt.
- Zellen zählen, Ausfüllen und alle Transformationen in den Kategorien Zahlenspalten, Datums- & Uhrzeitspalten und Strukturierte Spalten werden im Kapitel 11 behandelt.
- Spalte pivotisieren und Spalte entpivotisieren werden im Kapitel 14 behandelt.
- Format, Analysieren und Extrahieren (Kategorie Textspalte) werden im Kapitel 9 behandelt.

In diesen Kapiteln werden die Transformationen von Power Query separat hervorgehoben.

7.2.1. Datentyp anpassen

Die wichtigste Transformation ist die Anpassung des Datentyps. Der Datentyp bestimmt, wie die Daten interpretiert werden. Die Interpretation der Daten bestimmt, welche Operationen auf die Daten angewendet werden können.

! Wichtig

Die Datentypen müssen im Schritt Geänderter Typ (Windows) bzw. geänderter Spaltentyp (MacOS) angepasst werden.

Power Query zeigt neben den Spaltenüberschriften mit einem Symbol den Datentyp der Spalte an. Tabelle 7.1 zeigt die Symbole für die verschiedenen Datentypen. Zusätzlich wird der Datentyp im Menubalken neben dem Kommando Datentyp angezeigt. Die Datentypen sind in vier Kategorien unterteilt:

- Zahlen
- Datum und Uhrzeit
- Zeichenketten
- Wahrheitswerte

Im Gegensatz zu Excel behandelt Power Query Datum und Uhrzeit als eigenen Datentyp. Power Query muss diese Unterscheidung machen, weil in den Datenquellen diese Werte als Zeichenketten kodiert vorliegen. Dabei handelt es sich um *strukturierte Daten*. Power Query hat für diese Daten spezielle *Parser*, die die Zeichenketten aus der Datenquelle in Zahlenwerte für Excel umwandeln.

Tabelle 7.1.: Power Query Datentypen mit Symbolen

Datentyp	Symbol	Kategorie	Bemerkungen
Text	$A^{B}C$	Zeichenketten	
TRUE/FALSE	×,	Wahrheitswerte	
Dezimalzahl	1.2	Zahlen	
Ganze Zahl	1 ² 3	Zahlen	
Prozentzahl	%	Zahlen	Dezimalzahl mal 100
Währung	\$	Zahlen	Dezimalzahl mit Währungsformatierung

Datentyp	Symbol	Kategorie	Bemerkungen
Binärzahl	010 101	Zahlen	Ganzzahl
Zahl mit Gebietsschema	ABC 123	Zahlen	Nur MacOS; Das Gebietsschema legt den Tausender- und Dezimaltrenner fest
Datum/Uhrzeit	Ō	Datum und Uhrzeit	Zeitstempel als formatierte Dezimalzahl
Datum		Datum und Uhrzeit	Ganzzahl mit Datumsformatierung
Uhrzeit	(L)	Datum und Uhrzeit	Dezimalzahl zwischen 0 und 1 mit Zeitformatierung
Dauer	Č	Datum und Uhrzeit	Dezimalzahl als Zeitdauer
Datum/Uhrzeit/Zeit	ZZC C	Datum und Uhrzeit	Internationalisierter Zeitstempel als formatierte Dezimalzahl, die Zeitzone muss in den Werten kodiert sein, sonst wird +1 angeommen.

Der Datentyp einer Spalte lässt sich mit dem Kommando Datentyp anpassen. Wird auf das Kommando im Menubalken geklickt, öffnet sich eine Auswahlliste, aus der der richtige Datentyp ausgewählt werden kann (Abbildung 7.6). Wird der Datentyp geändert, dann erfolgt eine Abfrage (s. Abbildung 7.7), ob der Datentyp im aktuellen Arbeitsschritt erstzt werden soll (Aktuelle ersetzen) oder ob ein neuer Arbeitsschritt eingefügt werden soll (Neuen Schritt hinzufügen). Hier kann normalerweise Aktuelle ersetzen ausgewählt werden.

7.2.2. Spalten umbenennen

Falls einzelne Spaltenüberschriften einer Datenquelle unhantlich sind, lassen sich diese Spalten umbenennen. Dazu wird das Kommando Umbenennen verwendet. Das Kommando ist im Menuband Transformieren im Abschnitt Beliebige Spalte zu finden. Das Kommando öffnet einen Dialog, in dem der neue Name der Spalte eingegeben werden kann. Die Überschrift wird in einem eigenen Arbeitsschritt geändert.

Spalten lassen sich auch durch einen Doppelklick auf den Spaltennamen umbenennen.

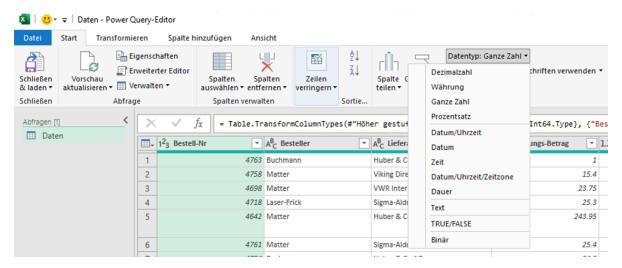


Abbildung 7.6.: Änderung des Datentyps einer Spalte



Abbildung 7.7.: Abfrage vor dem Ändern eines Datentyps

Die Spaltenüberschriften müssen für die Datenstruktur eindeutig sein. Wird eine Spaltenüberschrift doppelt verwendet, dann gibt Power Query eine Fehlermeldung aus.



Das Kommando Umbenennen sollte nur verwendet werden, wenn nur einzelne Spalten umbenannt werden müssen. Falls mehrere Spalten umbenannt werden müssen, sollte dieser Schritt als Transformation in Excel durchgeführt werden.

7.2.3. Spalten verschieben

In Excel können nur zusammenhängende Bereiche einer Tabelle mit einer Adressierung gemeinsam angesprochen werden. Diese Reihenfolge ist immer dann hilfreich, wenn Matrix-Operationen mit diesen Daten ausgeführt werden sollen. Falls die Spalten einer Datenquelle in einer ungünstigen Reihenfolge vorliegen, lassen diese sich mit einem Kommando unter Verschieben richtig positionieren.

7.2.4. Spalte teilen

Manche Daten fassen mehrere Werte in der einer Spalte zusammen. Falls die Werte als strukturierte Daten vorliegen, kann diese Spalte mit dem Kommando Spalte teilen in mehrere Spalten aufgeteilt werden. Dabei werden drei Varianten unterschieden.

- 1. Die Separator-Trennung verwendet Trennzeichen zum Abgrenzen der Werte.
- 2. Die Positions-Trennung trennt Festkodierungen mit festen Positionen in einzelne Spalten.
- 3. Die Längen-Trennung trennt Festkodierungen, deren Werte immer die gleiche Anzahl von Symbolen haben.

Die Separator-Trennung kann ein festes Trennzeichen oder ein Trennungsmuster verwenden. Die möglichen Trennungsmuster sind:

- Beim Wechsel zwischen kleinen und grossen Buchstaben.
- Beim Wechsel zwischen grossen und kleinen Buchstaben.
- Beim Wechsel zwischen Ziffern und Nicht-Ziffern.
- Beim Wechsel zwischen Nicht-Ziffern und Ziffern.

Beim trennen von Spalten dürfen nicht semistrukturierte Daten getrennt werden, weil diese nicht für alle Werte die gleiche Struktur haben.

7.3. Spalten auswählen oder entfernen

Zwei besondere Transformationen sind Spalten auswählen und Spalten entfernen. Diese Transformationen werden im Menuband Start angeboten. Mit diesen Transformationen lassen sich unbenötigte Spalten aus den Daten entfernen. Dadurch werden die Daten in Excel übersichtlicher.

Diese Funktion kommt bei Excel-Arbeitsmappen häufig zum Einsatz, wenn nach Ende der Daten zusätzliche leere Spalten folgen. Diese Spalten können mit Spalten entfernen entfernt werden.



Abbildung 7.8.: Power Query Kommandos zum auswählen oder entfernen von Spalten (Menu Start)

Beide Kommandos verändern die Datenstruktur, indem Spalten entfernt werden. Das Kommando Spalten auswählen lässt die ausgewählten Spalten in der Datenstruktur und entfernt alle anderen. Das Kommando Spalten entfernen entfernt nur die ausgewählten Spalten.



Achtung

Eine spezielle Variante von Spalten auswählen ist Kommando In Liste konvertieren im Menuband Transformieren. Dieses Kommando entfernt alle bis auf die ausgewählte Spalte. Dieses Kommando ändert jedoch die Spaltenüberschrift und sollte deshalb nicht verwendet werden.

7.4. Daten aktualisieren

Die Daten können sich aber ändern, z.B. weil ein Formular ausgefüllt wurde oder eine Datenbank aktualisiert wurde. Dadurch ändern sich die Daten in der Datenquelle. In solchen Fällen ist die Kopie in der Arbeitsmappe nicht mehr aktuell. Um die Daten zu aktualisieren, muss die Datenverbindung aktualisiert werden.

Dazu wird das Kommando Alle aktualisieren verwendet. Das Kommando ist im Menu Daten im Abschnitt Verbindungen zu finden. Das Kommando aktualisiert alle Datenverbindungen in der Arbeitsmappe entsprechend der Importspezifikation.



Abbildung 7.9.: Menubalken Daten - Kommando Alle aktualisieren

Es werden nur die importierten Daten aktualisiert. Wurde die Tabelle durch Formeln erweitert, werden diese Formeln *nicht* gelöscht, sondern auf die neuen Daten erweitert.

Wichtig

Die importierte Struktur darf **nicht verändert** werden. Das bedeutet, dass innerhalb der Struktur keine Spalten hinzugefügt oder gelöscht werden dürfen. Neue Spalten für Formeln müssen rechts von der importierten Datenstruktur der Tabelle hinzugefügt werden.

7.5. Datenverbindung anpassen

Gelegentlich ändert sich der Ort einer Datenquelle. Sehr häufig tritt diese Situation ein, wenn eine Datei in einen anderen Ordner verschoben wird. In solchen Fällen muss die Datenverbindung angepasst werden. Dazu muss in Power Query die Datenquelle angepasst werden.

- 1. Powerquery öffnen
- 2. Einstellungen im Schritt "Quelle" auswählen (s. Abbildung 7.10)
- 3. Pfad zur Datenquelle anpassen

Nachdem die richtige Datei ausgewählt wurde, sollte immer das Schema kontrolliert werden, ob die Änderung keine Auswirkungen auf das Schema hat. Wird dieser Schritt übersprungen, kann es zu Fehlern beim Import kommen und im schlimmsten Fall werden die Formeln in der Arbeitsmappe durch eine Fehlerfortpflanzung zerstört.

Abschliessend kann wieder das Kommando Schliessen & Landen ausgeführt werden.

Alle Datenverbindungen werden auch im Dialog **Abfragen & Verbindungen** (s. Abbildung 7.11) angezeigt. Der Dialog ist in der Gruppe **Verbindungen** im Abschnitt **Daten** zu finden. Der Dialog zeigt alle Datenverbindungen an, die in der Arbeitsmappe verwendet werden.

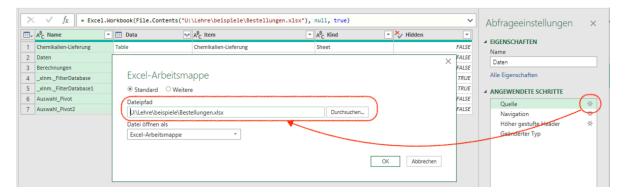


Abbildung 7.10.: Power Query: Quelle Anpassen (Ausschnitt)

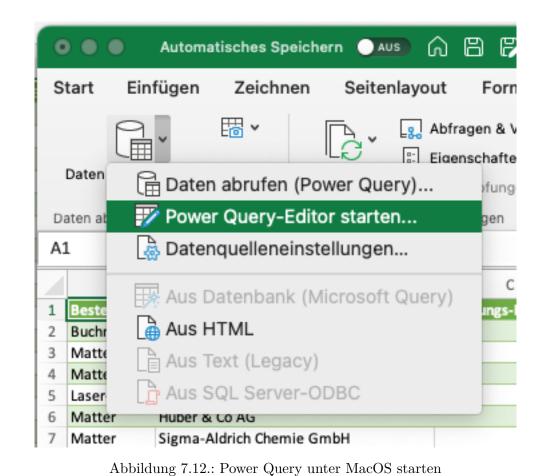


Abbildung 7.11.: Menubalken Daten - Abfragen & Verbindungen



⚠ MacOS vs. Windows

Unter Windows lassen sich die Datenquellen im Dialog Abfragen & Verbindungen anpassen. Die Windows Version von Excel von dort immer den Power Query Editor. Die Version für MacOS macht das nicht. Um Änderungen an einer Abfrage vorzunehmen, muss die Abfrage in Power Query über das Kommando Daten abrufen geöffnet werden (s. Abbildung 7.12).



Teil III. Mathematik der Daten

8. Variablen, Funktionen und Operatoren



Work in Progress

8.1. Variablen

Excel kennt keine strikte Unterscheidung zwischen Konstanten und Variablen. Grundsätzlich sind alle Werte in Excel Konstanten, weil direkt eingegebene oder durch eine Formel erzeugte Werte nicht durch andere Formel verändert werden können.



Warnung

In Excel lassen sich Zellen und Bereiche sperren. Damit lassen sich die Werte wie Konstanten behandeln. Diese Sperre ist Teil der Formatierung einer Zelle und muss für das gesamte Arbeitsblatt festgelegt werden. Die Sperre kann nicht für einzelne Zellen oder Bereiche festgelegt werden.

Die Sperre eines Arbeitsblatts wird im Menuband Überprüfen mit dem Kommando Blatt schützen aktiviert. Richten Sie bei diesem Kommando kein Passwort ein.

Ist ein Arbeitsblatt gesperrt, sind die Werte auf diesem Arbeitsblatt unveränderlich und verhalten sich wie Konstanten. Ist ein Arbeitsblatt nicht gesperrt, dann verhalten sich die Werte wie Variablen.

8.1.1. Dynamische Bereiche

Formeln können mehr als einen Wert verarbeiten und mehr als einen Ergebnis liefern. Solche Formeln müssen nur in die linke obere Ecke eines Bereichs eingetragen werden. Excel erkennt automatisch, dass die Formel auf einen Bereich angewendet werden soll und erzeugt die entsprechende Formel für alle Zellen des Bereichs. Das Ergebnis einer solchen Formel ist ein dynamischer Bereich.

Definition 8.1. Vektorisieren heisst das Erzeugen eines dynamischen Bereichs aus einem statischen Bereich.

Beispiel 8.1 (Vektorisieren eines statischen Bereichs).

=A1:A10

Im Gegensatz zu einem normalen Bereich, ist bei einem dynamischen Bereich nur die linke obere Zelle bekannt. Um trotzdem alle Zellen eines solchen Bereichs zu adressieren, wird die Gatter- (#) Notation verwendet.

Das Beispiel 8.1 erzeugt einen Bereich mit 10 Zellen. Die Formel wird in die linke obere Zelle des Bereichs eingetragen. Die Formel wird z.B. in die Zelle B1 eingetragen. Anschliessend können die Werte im Bereich B1:B10 über die Gatter-Notation addressiert werden: B1#.

Der Vorteil des Vektorisierens ist, dass der Bereich durch das Einfügen neuer Zeilen vergrössert werden kann, ohne dass die nachfolgenden vektorisierten Formeln angepasst werden müssen.



Praxis

Weil Tabellen automatisch vektorisiert werden, ist es einfacher Werte in einer Tabelle zu erfassen bzw. als Tabelle zu importieren (s. Kapitel 7) und anschliessend über die Tabellenadressierung auf die Werte zu verweisen.

8.1.2. Benannte Bereiche

Im Funktionsbalken wird ganz links die Adresse der aktuellen Zelle angezeigt. Wenn in dieses Feld geklickt wird, dann kann der Adresse ein Name zugewiesen werden. Auf diese Weise kann eine Zelle oder ein Bereich benannt werden. Der Name eines Bereichs darf nur einmal in einer Arbeitsmappe existieren. Dafür kann in jeder Zelle der Arbeitsmappe dieser Namen als Adresse verwendet werden. Dazu ist es nicht notwendig, den die genaue Arbeitsblatt- und Zellenadresse zu kennen.

Verweist ein benannter Bereich mit einer Zelle auf den Anfang eines dynamischen Bereichs, dann kann der gesamte dynamische Bereich mit der Gatter-Notation referenziert werden.

Auf diese Weise lassen sich oft adressierte Zelleen oder Bereiche benennen. Dadurch kann die Adresse für Verwendung in Formeln abstrahiert werden.

Die Verwendung von benannten Bereichen wird weiter vereinfacht, dass bei der Eingabe von Formeln die Namen der benannten Bereiche als Vorschlag angezeigt werden.

Beispiel 8.2 (Adressierung eines dynamischen benannten Bereichs).

=umsatz#

Benannte Zellen helfen, auf wichtige Zellen in einer Arbeitsmappe zuzugreifen.

i Merke

Ein benannter Bereich entspricht einer Variablen in anderen Programmiersprachen.

Wichtig

Eine benannte Zelle oder ein benannter Bereich sind immer absolut referenziert.

Hinweis

Wenn einem benannten Bereich Zellen hinzugefügt werden, dann wird der benannte Bereich um diese Zellen erweitert.

i Hinweis

Eine Tabelle ist ein spezieller benannter Bereich, der als Ganzes oder in Teilen referenziert werden kann (s. Abschnitt 6.4.2).

8.2. Funktionen

8.3. Operatoren

Excel hat eine Anzahl vordefinierter Operatoren. Nicht alle Excel Operatoren haben keine direkte Entsprechung als Funktion.

8.3.1. Substitution über Funktionspfade

8.3.2. Substitution mit LET()

8.4. Funktionsketten

• Funktionsketten mit LET

8.4.1. Leere Zellen in einer Funktionskette

Normalerweise werden leere Zellen als Ergebnis einer Funktion durch 0 ersetzt. Dieses Konvertierung findet erst bei der Darstellung des Ergebnisses statt. Innerhalb einer Funktionskette werden leere Zellen als leere Zellen weitergereicht, solange keine Aggregation

vorgenommen wird. Es ist deshalb möglich in einer Funktionskette eine Entscheidung mit ISTLEER() für den Fall einer leeren Zelle zu treffen.

Funktionen selbst definieren

LAMBDA Funktion

Strategie der Funktionsentwicklung - Funktion als Funktionspfad - Funktion als Funktionskette - Funktion als LAMBDA Funktion - Funktion als benanntes Element

9. Zeichenketten



lack Work in Progress

10. Aussagenlogik

Dieses Kapitel befasst sich mit der Anwendung der Aussagenlogik in Excel. Es werden die logischen Operatoren NICHT, UND, ODER und XODER behandelt. Auf dieser Grundlage werden die wichtigsten Vergleichsoperation zum Formulieren logischer Ausdrücke vorgestellt. Das Kapitel schliesst mit der Anwendung von Fallunterscheidungen.

Es werden die folgenden Funktionen behandelt: NICHT(), UND(), ODER(), XODER(), WENN(), WENNS(), ERSTERWERT(), WENNFEHLER(), FEHLER.TYP(), ISTFEHLER(), XVERWEIS(), XVERGLEICH(), IDENTISCH(), FILTER(), SORTIEREN(), SORTIERENNACH() und SPALTENWAHL()

10.1. Wahrheitswerte in Excel

Wie im Kapitel Datentypen bereits erwähnt, kennt Excel den Datentyp der Wahrheitswerte. Diese Wahrheitswerte können entweder den Wert WAHR oder FALSCH haben. Eine Operation, die Wahrheitswerte ergibt wird als logischer Ausdruck bezeichnet. Weil logische Ausdrücke für viele Funktionen und Operationen notwendig sind, wandelt Excel die Werte anderer Datentypen bei Bedarf um. Dabei gelten die folgenden Regeln:

- 0 und die leere Zelle entspricht dem Wert FALSCH.
- Alle Zahlen ungleich 0 entstprechen dem Wert WAHR.
- Alle Zeichenketten inklusive der leeren Zeichenkette entsprechen dem Wert WAHR.
- Fehlerwerte bleiben unverändert.

Werden Wahrheitswerte in mathematischen Operationen und Funktionen verwendet, dann konvertiert Excel den Wert FALSCH in 0 und den Wert WAHR in 1 um.

Werden Wahrheitswerte als Parameter an Zeichenkettenfunktionen übergeben, dann werden die Wahrheitswerte in die entsprechende Zeichenkette umgewandelt. Aus dem Wert WAHR wird also die Zeichenkette "WAHR" und aus dem Wert FALSCH wird die Zeichenkette "FALSCH".

⚠ Warnung

Wahrheitswerte werden in den verschiedenen Sprachversionen von Excel in der eingestellten Sprache angegeben. Beim Wechsel zwischen verschiedenen Excel-Sprachversionen werden die Wahrheitswerte automatisch korrekt angezeigt. Die Umwandlung in Zeichenketten erfolgt dann in der jeweiligen Sprache. Deshalb sollte die Verwendung von in Zeichenketten konvertierten Wahrheitswerten in nachgelagerten Funktionen vermieden werden.

10.2. Aussagenlogische Operationen

Für Wahrheitswerte existieren spezielle Operationen, um die Regeln der Aussagenlogik bzw. der Boole'schen Algebra abzubilden. Diese Operationen verknüpfen Wahrheitswerte und haben Wahrheitswerte als Ergebnis. Die vier Grundoperationen NICHT (\neg) , UND (\land) , ODER $(\lor$, "inklusives Oder") und XODER (\oplus) , "entweder-oder") sind in Excel als Funktionen verfügbar.

Die Funktion NICHT() wandelt einen Wert in den jeweils den anderen Wahrheitswert um. Falls anstelle eines Wahrheitswerts ein anderer Datentyp übergeben wurde, gelten die oben angegebenen Regeln für die Umwandlung.

Die Funktionen UND(), ODER() und XODER() sind Aggregatoren. Das bedeutet, dass Sie alle Werte in dem angegebenen Bereichen zusammenfassen. Das ist oft nicht das gewünschte Verhalten. Deshalb muss bei der Arbeit mit Vektoren auf die Bool'sche Arithmetik zurückgegriffen werden, um logische Ausdrücke richtig auszuwerten.

Zum Beispiel sollen für die folgenden Werte paarweise der logische Ausdruck $a \wedge b$ ausgewertet werden, so dass für alle Wertepaare der richtige Wahrheitswert ermittelt wird.

A	В
WAHR	FALSCH
FALSCH	WAHR
WAHR	WAHR
FALSCH	WAHR
WAHR	WAHR
FALSCH	FALSCH

Die Formel = UND(A1:A6; B1:B6) liefert den Wert FALSCH zurück, weil nicht alle Werte im gesamten Bereich von A1:A6 und B1:B6 gleich WAHR sind. Es gibt keine Funktion und keinen eigenen logischen Operator zur paarweisen logischen Verknüpfung dieser beiden Bereiche. Deshalb werden in Excel oft logische Ausdrücke in der Boole'schen Arithmethik eingesetzt, um

nur die Werte aus den gleichen Datensätzen miteinander zu vergleichen. Diese Schreibweise ist immer dann notwendig, wenn logische Ausdrücke sich auf die einzelnen Datensätze beziehen.

Die Formel = A1:A6 * B1:B6 hat die Werte {0;0;1;0;1;0} zum Ergebnis. Um Wahrheitswerte zuerhalten kann noch auf die Ungleichheit mit 0 geprüft werden. Dazu wird die Formel wie folgt ergänzt: = (A1:A6 * B1:B6) <> 0. Das Ergebnis ist nun {FALSCH; FALSCH; WAHR; FALSCH}. Dieser Schritt ist in der Praxis selten notwendig, weil für die meisten Operationen Zahlenwerte implizit als Wahrheitswerte behandelt werden.

Weil Excel alle Werte ungleich 0 als WAHR interpretiert, können die Operationen UND() mit * und die Operation ODER() mit + direkt ersetzt werden. Hierbei ist darauf zu achten, dass das nummerische Ergebnis dieser Addition oder Multiplikation ausschliesslich als Wahrheitswert von Bedeutung ist.

Die Operation XODER() entspricht der Ungleichheit <>. Dabei muss allerdings darauf geachtet werden, das dieser Vergleich als Ersatz für XODER() ausschliesslich für Wahrheitswerte bzw. 0 und 1 erlaubt ist. Für die oben gezeigten Werte ergibt die Formel = (A1:A5 <> B1:B5) die Werte {WAHR; WAHR; FALSCH; WAHR; FALSCH; FALSCH}. Das Ergebnis ist deshalb sichergestellt, weil alle Vergleichswerte 0 oder 1 sind. Werden jedoch auch andere Vergleichswerte zugelassen, dann liefert die Formel = (A1:A6 <> B1:B6) die Werte {WAHR; WAHR; WAHR; WAHR; WAHR}. Dieses Verhalten zeigt das folgende Beispiel.

1 0 0 2 1 2 0 1 3 2 0 0	A	В
1 2 0 1 3 2	1	0
0 1 3 2	0	2
3 2	1	2
	0	1
0 0	3	2
	0	0

Damit das Richtige Ergebnis erzeugt wird, müssen die Werte in Wahrheitswerte konvertiert werden. Dazu muss die Formel durch Vergleiche ungleich 0 erweitert werden:

```
= (A1:A6 <> 0) <> (B1:B6 <> 0)
```

Diese Formel liefert die gewünschten Werte {WAHR; WAHR; FALSCH; WAHR; FALSCH; FALSCH}.

Die folgende Tabelle zeigt logischen Operatoren und die zugehörigen Terme für die Boole'sche Arithmetik.

Operator	Boole'sche Operation	Vereinfachter Operator
	1 - a	NICHT(a)
\wedge	a * b	a * b
V	a + b - a * b	a + b
\oplus	a + b - 2 * a * b oder (a - b) ^ 2	(a <> b)

10.3. Vergleiche

Eine besondere Art von logischen Ausdrücken sind *Vergleiche*. Ein Vergleich prüft das Verhältnis zweier Werte zueinander. Excel kennt die üblichen Vergleichsoperatoren, die jeweils einen Wahrheitswert zurückliefern.

In Excel werden die Vergleichsoperatoren wie folgt geschrieben:

- > (grösser als)
- < (kleiner als)
- >= (grösser oder gleich)
- <= (kleiner oder gleich)
- = (gleich)
- <> (ungleich)

Excel's Vergleichsoperatoren sind Datentypen sensitiv. Das bedeutet, dass die Operatoren Datentypen vor dem Vergleich *nicht* angleichen. Der folgende Vergleich ergibt also FALSCH.

Merke

Ziffern sind keine Zahlen!

Weil die Operationen *, + und – normalerweise vor den Vergleichsoperatoren ausgeführt werden, müssen alle Vergleiche eines logischen Ausdrucks für die Boole'sche Arithmetik in Klammern gesetzt werden.

10.3.1. Der ∈-Operator mit XVERWEIS

Mithilfe der Funktion XVERGLEICH() kann der ∈-Operator aus der *Mengenlehre* in Excel für logische Ausdrücke bereitgestellt werden. Mit dieser Funktion XVERWEIS() können sowohl der ∈ als auch der ∉-Operator mit XVERWEIS() abgebildet werden.

Für die folgenden Beispiele verwenden wir die Werte:

A	В
Suchkriterium	Suchbereich
4	1
7	3
	4
	8

←-Operator: XVERWEIS(A2:A3; B2:B5; B2:B5 = B2:B5; FALSCH)
 ∉-Operator: XVERWEIS(A2:A3; B2:B5; B2:B5 <> B2:B5; WAHR)

Der Trick besteht darin, dass die *Rückgabematrix* durch einen Vergleich aus dem Suchbereich erzeugt wird. Dadurch wird die Rückgabematrix mit den gleichen Wahrheitswerten für alle Werte im Suchbereich gefüllt. Die erste Formel ergibt deshalb {WAHR; FALSCH} und die zweite Formel {FALSCH; WAHR}, weil der Wert 4 im Suchbereich vorkommt und der Wert 7 nicht. Diese Werte können direkt in logischen Ausdrücken verwendet werden.

10.3.2. Zeichenketten vergleichen

Die Vergleichsoperatoren zeigen die Unterschiede zweier Zeichenketten bezüglich der alphabetischen Sortierung an. Die "kleinste" Zeichenkette ist die leere Zeichenkette. Gross- und Kleinschreibung wird bei Zeichenkettenvergleichen nicht unterschieden.

Weil Excel für Vergleiche die nicht-druckbaren Zeichen mit Ausnahme des Leerzeichens und des Tabulators ignoriert, gibt der Vergleichsoperator = WAHR auch für Zeichenketten zurück, die unterschiedliche nicht-druckbare Zeichen enthalten. Das gleiche Problem entsteht beim Vergleich von unterschiedlicher Gross- und Kleinschreibung. Um auch diese Unterschiede zu erkennen, müssen wir die Funktion IDENTISCH() verwenden. Diese Funktion vergleicht die Zeichenketten Zeichen für Zeichen und liefert nur dann WAHR zurück, wenn die Zeichenketten exakt gleich sind.

Beispiel 10.1. ## Anwendung der IDENTISCH() Funktion als Vergleichsoperator.

```
= WENN(IDENTISCH("A"; "a"); "Gleich"; "Ungleich")
```

10.4. Komplexe logische Ausdrücke und Datenstrukturen

Excel hat zwar Funktionen für die logischen Operatoren *Und* (UND()), Oder (ODER()) und *Exklusives Oder* (XODER()), diese Funktionen haben aber den Nachteil, dass sie nur Werte zusammenfassen können. In der Praxis werden jedoch oft *Datenstrukturen* als Variablen für

logische Ausdrücke verwendet. Diese sollen durch die logischen Operatoren verknüpft und nicht zusammengefasst werden.

Um logische Verknüfungen für Datenstrukturen zu realisieren, müssen die logischen Operatoren mit der Boole'schen Arithmetik umgesetzt werden. Dazu werden die logischen Operatoren durch die entsprechenden arithmetischen Operatoren ersetzt.

Beispiel 10.2. Für die folgenden Werte soll der folgende logische Ausdruck geprüft werden.

$a \wedge b \vee$	(c <	10)	$\wedge d$

A	В	С	D
WAHR	5	21	17
FALSCH	3	5	1
WAHR	0	10	2
FALSCH	1	11	3

Für diesen Ausdruck werden die Werte {WAHR; WAHR; FALSCH; FALSCH} erwartet.

Die naive Umsetzung =ODER(UND(A1:A4;B1:B4);UND(C1:C4;D1:D4)) hat als Ergebnis den Wert WAHR.

Die Formel = A1:A4 * B1:B4 + (C1:C4 < 10) * D1:D4 liefert die Werte {5; 1; 0; 0} diese lassen sich durch einen Vergleich in die entsprechenden Wahrheitswerte konvertieren. Diese Konversion ist nur notwendig, wenn der Ausdruck nicht als logischer Ausdruck an eine Funktion übergeben wird.

$$= (A1:A4 * B1:B4 + (C1:C4 < 10) * D1:D4) <> 0$$

Diese Formel ergibt die erwarteten Werte {WAHR; WAHR; FALSCH; FALSCH}.

Die Ausnahme von dieser Regel ist Verwendung der logischen Funktionen als Tabellenfunktion. In diesem Fall können die Werte zeilenweise auch von den logischen Funktionen verarbeitet werden.

Beispiel 10.3. Werden die Werte aus Beispiel 10.2 als Tabelle konvertiert und mit BeispielTabelle benannt, dann können die logischen Funktionen in einer gleich langen Tabelle mit der gleichen Startzeile auch als Tabellenfunktionen verwendet werden. In diesem Fall werden die Werte zeilenweise verarbeitet.

```
=ODER(
     UND(BeispielTabelle[@a];BeispielTabelle[@b]);
     UND(BeispielTabelle[@c] < 10; BeispielTabelle[@d])
)</pre>
```

Diese Formel ergibt die erwarteten Werte {WAHR; WAHR; FALSCH; FALSCH}.

10.5. Fälle unterscheiden

Logische Ausdrücke eigenen sich besonders gut, um Fallunterscheidungen zu formulieren, weil ein logischer Ausdruck immer nur zwei Werte als Ergebnis haben kann. Es gibt also für jeden logischen Ausdruck immer nur zwei unterscheidbare Fälle.

Excel hat zwei zentrale Funktionen für Fallunterscheidungen: WENN() und WENNS(). Die Funktion WENN() ist eine einfache Unterscheidung, die Funktion WENNS() unterstützt mehrfache Unterscheidungen. In anderen Programmiersprachen wird in diesem Zusammenhang auch von Verzweigungen gesprochen.

10.5.1. WENN

Die Funktion WENN() ist eine *einfache* Fallunterscheidung. Einfach bedeutet hier, dass die beiden Fälles *eines* logischen Ausdrucks unterschieden werden. Entsprechend hat die Funktion WENN() drei Parameter:

- 1. Der auszuwertende logische Ausdruck.
- 2. Das Ergebnis falls der logische Ausdruck WAHR ergibt.
- 3. Das Ergebnis falls der logische Ausdruck FALSCH ergibt.

Das Ergebnis für den Fall, dass der logische Ausdruck FALSCH ergibt, ist optional. Fehlt dieser Parameter, dann wird der Wert FALSCH zurückgegeben.

Das Verhalten dieser Funktion lässt sich mit den Wahrheitswerten als logischer Ausdruck direkt überprüfen:

```
= WENN(WAHR; "Guten Tag"; "Auf Wiedersehen")
```

Weil der logische Ausdruck in diesem Fall WAHR ist, wird der zweite Parameter als Ergebnis zurückgegeben. Die Formel gibt also den Wert "Guten Tag" zurück.

Wird der logische Ausdruck auf FALSCH geändert, dann liefert die Formel den Wert "Auf Wiedersehen".

```
= WENN(FALSCH; "Guten Tag"; "Auf Wiedersehen")
```

Lassen wir den dritten Parameter weg, dann wird der Wert FALSCH zurückgegeben.

```
= WENN(FALSCH; "Guten Tag")
```

Ausser der Fallunterscheidung hat WENN() keine weiteren Eigenschaften. Deshalb wird diese Funktion in der Praxis oft mit anderen Funktionen kombiniert. Das kann mit der Funktion WENN() selbst geschehen. In diesem Fall wird von geschachtelten Fallunterscheidungen gesprochen.

Zur Veranschaulichung dient das folgende Beispiel:

	A
1	4
2	7

Eine Fallunterscheidung soll prüfen, ob die Werte in A1:A2 Werte gleich 1, 3, 4 oder 8 sind. Falls das der Fall ist, soll der zugehörige Zahlwert als Zeichenkette ausgegeben werden. Falls das nicht der Fall ist, soll der Wert Ungültig zurückgegeben werden. Als geschachtelte WENN()-Funktion lässt sich diese Fallunterscheidung wie folgt formulieren:

```
= WENN(A1 = 1; "Eins";
    WENN(A1 = 3; "Drei";
    WENN(A1 = 4; "Vier";
    WENN(A1 = 8; "Acht";
    "Ungültig"))))
```

Eine solche geschachtelte Fallunterscheidung wird als Entscheidungsbaum bezeichnet.

10.5.2. WENNS

Die Funktion WENN() ist eine einfache Fallunterscheidung. In vielen Excel-Arbeitsmappen existieren geschachtelte Aufrufe von WENN()-Funktionen. Diese Aufrufe machen die Formeln nicht nur schwer lesbar, sondern auch fehleranfällig und ineffizient. Deshalb sollten geschachtelte Fallunterscheidungen unbedingt vermieden werden. Mit der Funktion WENNS() lassen sich geschachtelte Fallunterscheidungen vermeiden, indem alle Fallunterscheidungen in einem einzigen Funktionsaufruf zusammengefasst werden.

i Merke

Geschachtelte Fallunterscheidungen mit WENN() unbedingt vermeiden!

Die Funktion WENNS() erwartet Parameterpaare, bestehend aus einem logischen Ausdruck und dem Ergebnis, falls dieser logische Ausdruck WAHR ergibt. Die Funktion kann bis zu 127 Parameterpaare verarbeiten, so dass sich auch sehr komplexe Fallunterscheidungen mit dieser Funktion abbilden lassen.

Beispiel 10.4. Das folgende Beispiel zeigt die Verwendung der Funktion WENNS() für die geschachtelte Fallunterscheidung aus dem Abschnitt WENN.

Das Beispiel bildet aber noch nicht die vollständige Fallunterscheidung ab. Es fehlt noch der Fall, dass keiner der logischen Ausdrücke WAHR ergibt. Leider kann WENNS() ausschliesslich logische Ausdrücke mit ihren Ergebnissen verbinden.

Merke

WENNS() kann nur logische Ausdrücke mit ihren WAHR-Ergebnissen verbinden.

Anders als bei WENN() gibt es keine direkte Möglichkeit, ein Ergebnis festzulegen, falls alle logische Ausdrücke FALSCH ergeben. Um ein solches Verhalten zu erzeugen, wird ausgenutzt, dass die Funktion WENNS() immer einen wahren logischen Ausdruck mit einem Ergebnis verknüpft. Weil die logischen Ausdrücke in der Reihenfolge ausgewertet werden, wie sie in der Funktion angegeben sind, muss der letzte logische Ausdruck alle Fälle abdecken, die von keinem anderen der vorangegangen logischen Ausdrücke akzeptiert wurden. Der einfachste logische Ausdruck, der immer wahr ist, ist der Wahrheitswert WAHR. Deshalb wird dieser Wert als letzter logischer Ausdruck für WENNS() verwendet.

Mit diesem Wissen lässt sich das Beispiel mit WENNS() vervollständigen:

Diese Formel prüft die Werte in A2: A3 auf Gleichheit mit den Werten 1, 3, 4 und 8. Für diese Zahlen wird die zugehörige Zahlwert als Zeichenkette ausgegeben. Falls keiner dieser Werte gefunden wird, wird der Wert Ungültig zurückgegeben.

Die Fallunterscheidung mit WENNS() endet beim ersten logischen Ausdruck, der WAHR ergibt. Die Funktion prüft der Reihe nach alle angegebenen logischen Ausdrücke. Sobald einer dieser Ausdrücke WAHR ist, wird der zugehörige Ergebniswert ausgegeben und die Funktion wird beendet. Diese Eigenschaft begründet, dass die logischen Ausdrück nur die Fälle prüfen müssen, die von den vorangegangenen logischen Ausdrücken nicht abgedeckt wurden.

Beispiel 10.5 (Fallunterscheidung mit WENNS() vereinfachen). Gegeben ist die Formel mit geschachtelten Entscheidungen.

Diese Formel ist aus zwei Gründen übermässig komplex.

- 1. Die Fallunterscheidung mit WENN() ist geschachtelt.
- 2. Es existieren redundante Fallunterscheidungen.

Bevor die Fallunterscheidung mit WENNS() vereinfacht wird, werden die redundanten Fallunterscheidungen entfernt. Das betrifft die zweite (J2 > L2) und die vierte Fallunterscheidung (J2 > N2). Im jeweiligen FALSCH-Fall wird der gegenteilige logische Ausdruck geprüft. Das ist in diesem Fall unnötig, weil die äussere Fallunterscheidung diesen Fall bereits abdeckt. Werden die redundanten logischen Ausdrücke und unnötige Klammern entfernt, dann ergibt sich die folgende wesentlich einfachere Formel.

```
K2+(M2-K2)/(N2-L2)*(J2-L2);
M2+(O2-M2)/(O2-N2)*(J2-N2)
)
);
100)
```

Die äusserste Fallunterscheidung hat für den Fall WAHR eine geschachtelte WENN()-Funktion und im Fall FALSCH ein einfaches Ergebnis. Das ist für WENNS() unhandlich, so dass die äusserste Fallunterscheidung durch Umkehrung des logischen Ausdrucks umgestellt wird.

Nun lassen sich die vier unterschiedlichen Fälle gut erkennen und mit WENNS() abbilden. Daraus ergibt sich die folgende Formel.

```
= WENNS(J2<02; 100;

J2>L2; 0;

(J2>N2)*(J2>=02); K2+(M2-K2)/(N2-L2)*(J2-L2);

WAHR; M2+(O2-M2)/(O2-N2)*(J2-N2)

)
```

Diese Formel ist wesentlich einfacher zu lesen und zu verstehen. Beim Durchgehen der Fälle fällt auf, dass ein Teilausdruck des dritten Falls das Gegenteil des ersten Falls ist. Diese Bedingung wurde bereits im ersten Fall geprüft und würde sie nicht gelten, dann wäre die Formel bereits beendet worden. Deshalb können bereits geprüfte Teilausdrücke in den nachfolgenden Ausdrücken weggefallen. Dadurch wird nicht nur die Verschachtelung, sondern auch die Komplexität der logischen Ausdrücke vereinfacht.

```
= WENNS(J2<02; 100;

J2>L2; 0;

J2>N2; K2+(M2-K2)/(N2-L2)*(J2-L2);

WAHR; M2+(O2-M2)/(O2-N2)*(J2-N2)
```

Diese Formel hat noch den Makel, dass der letzte Fall WAHR keine Konstante abbildet. Besser wäre es, wenn der zweite und der letzte Fall vertauscht wären, so dass der Wert 0 der letzte Wert ist. Dazu müssen die logischen Ausdrücke umorganisiert werden. Bei der Umorganisation ist die Reihenfolge der logischen Ausdrücke zu beachten: Die letzten beiden Fälle sind nicht umabhängig vom logischen Ausdruck J2>L2. Beim Umorganisieren darf diese Abhängigkeit nicht verloren gehen.

Diese Formel ist deutlich einfacher und weniger Fehleranfällig als die ursprüngliche Formel mit geschachtelten WENN()-Funktionen. Es lassen sich auch weitere Fälle hinzufügen, ohne dass die Formel komplexer wird. Dabei ist zu beachten, dass diese Fälle *vor* dem Fall WAHR angegeben werden müssen.

10.5.3. Nicht erreichbare Entscheidungen

Ein besonderes Problem sind Entscheidungen, die zwar definiert aber nie erreicht werden können. Solche Entscheidungen sind immer redundant. Eine nicht erreichbare Entscheidung kann nur dann auftreten, wenn eine vorangegangene Entscheidung bereits den geprüften Fall abdeckt. Ergibt ein solcher logischer Ausdruck Falsch, dann wird eine spätere Entscheidung für den gleichen Fall im Falsch-Zweig des Entscheidungsbaums ebenfalls Falsch ergeben. Der Wahr-Zweig dieser Entscheidung kann damit nie erreicht werden.

Beispiel 10.6 (Nicht erreichbare Entscheidung).

In Beispiel 10.6 kann nie das Ergebnis "Gut" erzeugt werden, weil der zweite logische Ausdruck (A1 > 3) alle Werte "maskiert", die durch den dritten logischen Ausdruck (A1 > 4) als "Gut" markiert werden müssten. "Ungenügend" würde trotzdem angezeigt werden, wenn der Wert in A1 entweder 1, 2 oder 3 ist.

Wichtig

Eine nicht erreichbare Entscheidung ist kein technischer Fehler, sondern ein logischer Fehler.

Im Beispiel 10.6 kann die Entscheidung A1 > 4 nicht erreicht werden, weil das vorherige und allgemeinere Kriterium A1 > 3 für die gleichen Werte zutrifft.

i Merke

Es müssen immer die spezielleren Kriterien vor den allgemeineren Kriterien geprüft werden.

Nicht erreichbare Entscheidungen lassen sich durch das Formale prüfen der logischen Ausdrücke leicht erkennen. Dazu werden logischen Ausdrücke und die zugehörigen Wertebereiche für den Wahr- und Falsch-Fall untereinander aufgeschrieben. Ein logischer Ausdruck kann einen Wertebereich nur dann abdecken, wenn dieser eine Teilmenge des Wertebereichs der aktuellen logischen Verzweigung ist.

Tabelle 10.7.: Formale Prüfung der logischen Ausdrücke aus Beispiel 10.6

Rang	logischer Ausdruck	Wahr-Fall	Falsch-Fall
1	A1 > 5	A1 > 5	A1 <= 5
2	A1 > 3	A1 > 3	A1 <= 3
3	A1 > 4	A1 > 4	A1 <= 4
4	A1 <= 3	A1 <= 3	A1 > 3

Weil die Funktion WENNS() verwendet wird, ist der Wertebereich für einen logischen Ausdruck durch die Falsch-Fälle der logischen Ausdrücke mit niedrigerem Rang abgedeckt.

Für Rang 3 muss wegen dieser Tabelle der logischen Ausdruck in Formel 10.1 gelten. Dieser Ausdruck kann jedoch nie Wahr ergeben, weil der geiche Wert in Variable A1 nicht kleiner oder gleich 3 und gleichzeitig grösser als 4 sein kann.

$$(A1 \le 5) \land \quad (A1 \le 3) \land (A1 > 4)$$

 $\Leftrightarrow \qquad (A1 \le 3) \land (A1 > 4)$
 $\Leftrightarrow \qquad \text{Falsch}$ (10.1)

10.5.4. ERSTERWERT

Die Funktion ERSTERWERT () bildet einen Spezialfall von WENNS () ab: Es wird bei allen logischen Ausdrücken ein Vergleich auf Gleichheit des Suchkriteriums mit verschiedenen Referenzwerten

durchgeführt. In diesem Fall können die logischen Ausdrücke mit ERSTERWERT() stark vereinfacht werden. Das lässt sich am ersten Beispiel im Abschnitt WENNS veranschaulichen.

Weil alle logischen Ausdrücke die Gleichheit über den gleichen Adressbereich prüfen, kann die Operation mit der Funktion ERSTERWERT() vereinfacht wie folgt werden.

```
= ERSTERWERT(A2:A3;
1; "Eins";
3; "Drei";
4; "Vier";
8; "Acht";
"Ungültig")
```

10.5.5. XVERWEIS zur Fallunterscheidung

Die Funktion XVERWEIS() ist als Excels Version des ∈-Operators bereits bekannt. Die Funktion kann auch als Alternative zur Funktion ERSTERWERT() verwendet werden. In diesem Fall werden als Rückgabematrix keine Wahrheitswerte, sondern die Ergebnisse der Fallunterscheidung angegeben.

Der Vorteil dieser Anwendung ist, dass die Fallunterscheidung nicht mehr auf die Anzahl der Parameterpaare beschränkt ist und die Parameterpaare zum Zeitpunkt der Formelerstellung auch nicht bekannt sein müssen.

Das folgende Beispiel zeigt die Umsetzung des Beispiels aus dem Abschnitt ERSTERWERT mit XVERWEIS(). Dazu wird zuerst eine Tabelle mit den Vergleichswerten und den zugehörigen Ergebnissen erstellt.

$\overline{\mathrm{C}}$	D
1	Eins
3	Drei
4	Vier
8	Acht

Diese Referenztabelle stell in Spalte C die sog. Suchmatrix und in Spalte D die sog. Rückgabematrix bereit. Mit diesen Werten lässt sich die Fallunterscheidung wie folgt abbilden.

```
= XVERWEIS(A2:A3; C2:D5; D2:D5; "Ungültig")
```

Ein weiterer Vorteil von XVERGLEICH() gegenüber ERSTERWERT() sind Vergleiche mit den Operatoren =, <= oder >=. Diese Vergleiche lassen sich über den fünften Parameter von XVERWEIS() konfigurieren. Dabei steht der Wert 0 für die Gleichheit, der Wert -1 für kleiner oder gleich und der Wert 1 für grösser oder gleich. Die Vergleiche sind immer so organisiert, dass der linke Operand dem Suchkriterium entspricht und der rechte Operand dem Wert in der Suchmatrix. Bei einem Treffer wird der Wert aus der Rückgabematrix zurückgegeben. Gibt es keinen Treffer für den Vergleich wird der Wert aus dem vierten Parameter wenn_nicht_gefunden geliefert.

10.5.6. Anwendungshilfe für Fallunterscheidungen

Die Anwendung der verschiedenen Fallunterscheidungsfunktionen hängt von verschiedenen Kriterien ab. Diese sind hier zusammengefasst:

Die Funktion WENN() wird immer dann eingesetzt, wenn ein logischer Ausdruck geprüft werden muss und nur die beiden Fälle dieses Ausdrucks unterschieden werden müssen.

Die Funktion WENNS() wird immer dann eingesetzt werden, wenn mehrere logische Ausdrücke geprüft werden müssen. Die logischen Ausdrücke können dabei beliebig komplex sein und sich auf verschiedene Daten und Bereiche beziehen.

Die Funktion ERSTERWERT() wird immer dann eingesetzt, wenn die Gleichheit des Suchkriteriums mit wenigen Referenzwerten überprüft werden soll. Die Suchkriterien sind für alle Vergleiche identisch.

Die Funktion XVERWEIS() wird immer dann eingesetzt, wenn ein Vergleich auf Gleichheit, Kleiner-oder-Gleich oder Grösser-oder-Gleich durchgeführt werden muss. Die Suchkriterien und die Vergleichsoperatoren sind für alle Vergleiche identisch.

Die Funktion XVERWEIS() muss anstatt von ERSTERWERT() verwendet werden, wenn die Referenzwerte des Vergleichs zum Zeitpunkt der Formelerstellung noch nicht bekannt sind oder leicht änderbar bleiben sollen.

10.6. Filtern

Excel bietet die Funktion FILTER() zum Filtern von Daten. Diese Funktion erzeugt einen Ergebnisbereich mit den Werten, die durch den angegebenen logischen Ausdruck ausgewählt wurden.

⚠ Warnung

In Excel können in Tabellen und Pivot-Tabellen für einzelne Vektoren Werte für die Darstellung "gefiltert" werden. Dabei verwendet Excel nicht die Filter Funktion, sondern blendet einzelne Datensätze aus. Dadurch können die Ergebnisse nachgereihter Operationen nicht mit den dargestellten Werten zusammenpassen, weil nicht-dargestellte Werte weiterhin Teil der Daten sind und bei Berechnungen weiterhin mitberücksichtigt werden.

Das Ergebnis der Funktion FILTER() ist ein dynamischer Bereich mit den ausgewählten Werten. Im Gegensatz zu Tabellen-Filter sind die nicht dargestellten Werte nicht mehr Teil der Daten. Deshalb sind die Ergebnisse der FILTER()-Funktion konsistent mit den Ergebnissen der nachgereihten Operationen.

Definition 10.1. Excels FILTER()-Funktion wählt aus einem Vektor die Werte aus, für die ein **Auswahlvektor** den Wert WAHR oder einen Wert, der dem logischen WAHR entspricht.

Beispiel 10.7 (Filtern mit Excel). Gegeben sind die folgende Werte in den Spalten A und B. Die Filter()-Funktion steht an Adresse D2.

	A	В	С	D	
1	Basel	WAHR		=FILTER(A1:A5;	B1:B5)
2	Genf	FALSCH		Basel	
3	Lugano	FALSCH		Zug	
4	Zug	WAHR		Zürich	
5	Zürich	WAHR			

Weil die Werte in Spalte B vom Datentyp Wahrheitswert sind, kann > dieser Vektor zur Auswahl der Städtenamen in Spalte A verwendet werden.

Anstelle eines Vektors mit Wahrheitswerten wird meistens ein Vergleich als zweiter Parameter übergeben. Dieser Vergleich muss einen Vektor erzeugen, der genauso lang ist, wie der Vektor im ersten Parameter. Solche Vektoren werden dynamisch erzeugt, indem ein Vergleich entweder den Vektor selbst oder einen benachbarten Vektor verwendet.

	A	В	С	D
1	Basel	deutsch		=FILTER(A1:A5; B1:B5 = "deutsch")
2	Genf	französisch		Basel
3	Lugano	italienisch		Zug

A	В	С	D
Zug Zürich	deutsch deutsch		Zürich

10.6.1. Excel Filter und logische Operationen

Weil die FILTER()-Funktion immer über Vektoren arbeitet, können die logischen Funktionen nicht verwendet werden, weil sie keine Vektoren erzeugen. Deshalb *muss* der logische Ausdruck des Filters als Boole'sche Arithmetik formuliert werden (Kapitel 10.4).

Hinweis

Die Funktion NICHT() ist kein Aggregator und kann mit der FILTER()-Funktion kombiniert werden.

Um mit komplexen logischen Ausdrücken in Filtern zu verwenden, müssen wir die logischen Operatoren durch ihre arithmetische Schreibweise ersetzen.

Tabelle 10.11.: Beispiel eines komplexen logischen Ausdrucks mit FILTER()

	A	В	C	D	E	
1	Name	Sprache	Einwohner:innen		\overline{Formel}	
2	Basel	deutsch	173863		=FILTER(A2:A6;(B2	:B6="deut
					> 100000))	
3	Genf	französisch	203856		Basel	
4	Lugano	italienisch	62315		Zürich	
5	Zug	deutsch	30934			
6	Zürich	deutsch	421878			

10.7. Selektieren

Sehr häufig liegen umfangreiche Daten mit vielen Vektoren vor. Soll sich eine Analyse auf einzelne Vektoren beschränken, dann sollen, analog zum Filtern von Datensätzen, nur diese Vektoren ausgewählt werden.

Definition 10.2. Das Filtern von Vektoren wird als **selektieren** bezeichnet.

Weil die Vektoren einer Stichprobe in der Regel benannt sind, werden Vektoren über ihre Namen selektiert.

i Merke

Die Vektorennamen einer Stichprobe haben besondere Eigenschaften:

- 1. Vektorennamen sind immer von Datentyp Zeichenkette.
- 2. Vektorennamen einer Stichprobe bilden einen Vektor.
- 3. Die Vektorennamen einer Stichprobe sind eindeutig

Die dritte Eigenschaft ist nicht ganz offensichtlich, denn in einer manuell eingegebenen zweidimensionalen Struktur kann eine Überschrift mehrfach verwendet werden. Sobald eine solche Struktur in eine Excel-Tabelle umgewandelt wird, erzwingt Excel eindeutige Vektorennamen.

Aus diesen Eigenschaften folgt, dass die Auswahl von Vektoren durch die Eigenschaften von Zeichenketten unterstützt wird. Wir können zur Auswahl die folgenden Operationen verwenden:

- Identischer Vektorname
- Vektorname beginnt mit einer bestimmten Zeichenkette
- Vektorname endet mit einer bestimmten Zeichenkette
- Vektorname enthält an einer beliebigen Position eine bestimme Zeichenkette

Diese Operationen lassen sich als logische Ausdrücke formulieren, wodurch sich komplexere Selektoren umsetzen lassen.

Der einfachste Weg zum Vektoren adressieren ist die Verwendung des Vektornamens über die Tabellenadressierung. Dabei wird der Vektorname in eckige Klammern gesetzt (s. Abschnitt 6.4.2). Die Tabellenadressierung ist jedoch auf vollständige Namen und auf zusammenhängende Bereiche beschränkt.

Die Verwendung einer Selektor-Funktion zur Auswahl von Vektoren ist nicht auf Tabellen beschränkt, sondern kann mit beliebigen tabellarischen Strukturen angewendet werden. In Excel wird eine Selektor Funktion durch die Funktionskette eines Filters mit der Funktion SPALTENWAHL() erreicht.

Neben der Tabellenadressierung bietet Excel die Funktion SPALTENWAHL(), um Vektoren aus einem Bereich auszulesen. Diese Funktion benötigt aber die Position der gewünschten Spalte, denn oft sollen aber Vektoren über Namen oder Namensteile ausgewählt werden. Dafür kommt die Funktion XVERGLEICH() zur Anwendung. Mit XVERGLEICH() erhalten wir die Position eines gesuchten Werts in einem angegebenen Bereich.

Die Idee hinter dem hier beschriebenen Ansatz ist, dass wir herausfinden, wo unser gewünschter Vektor in der Stichprobe steht. Anschliessend wählen wir alle Werte an dieser Position mit der Funktion Spaltenwahl() aus.

i Hinweis

Die Funktion XVERGLEICH() ähnelt der Funktion XVERWEIS() indem wir einen Wert in einem Vektor suchen können. Anstelle eines Referenzwerts aus einem anderen Vektor liefert XVERGLEICH() nur die Position des gesuchten Werts zurück. Falls ein Wert mehrfach vorkommt, dann gibt die Funktion nur die erste Position zurück.

Der Funktion XVERGLEICH() können mehrere Suchwerte übergeben werden, für welche die Positionen bestimmt werden.

4	Α	В	С	D	E	F	G	Н	ı	
1	foo	bar	baz			bar	foo			
2						=SPALTENWAHL(A3:C4; XVERGLEICH(F1:G1; A1:C1))				
3	1	3	5			3	1			
4	2	4	6			4	2			
_										

Abbildung 10.1.: Beispiel für Vektorenselektion

Der Algorithmus zum Selektieren von Vektoren ist durch die folgenden Schritte definiert:

- 1. Wir *vektorisieren* nur die Vektornamen auf einem neuen Arbeitsblatt mit der Identitätsfunktion ab Adresse B1. Zur Veranschaulichung nenne ich dieses Arbeitsblatt Stichprobendaten.
- 2. Wir *vektorisieren alle* Stichprobendaten mittels der Identitätsfunktion auf dem gleichen Arbeitsblatt ab Adresse B2.
- 3. Auf einem neuen Arbeitsblatt geben wir in der ersten Zeile die Vektorennamen ab Adresse A1 ein, die wir auswählen möchten. In diesem Beispiel wird angenommen, dass 3 Vektoren ausgewählt werden sollen.
- 4. Wir wählen die einzelnen Vektoren mit der folgenden Formel an der Adresse A2 aus.

Beispiel 10.8 (Selektion von Vektoren in Excel).

```
=SPALTENWAHL(Stichprobendaten!$B$2#;

XVERGLEICH(A1:C1; Stichprobendaten!$B$1#))
```

Der Vorteil dieser Strategie ist, dass die Selektion individuelle, nicht-zusammenhängende Vektoren selektieren kann und nicht auf Tabellen beschränkt ist.

10.8. Sortieren

Excel kennt zwei Funktionen zum Sortieren:

- SORTIEREN()
- SORTIERENNACH()

Die Funktion SORTIEREN() sortiert einen Bereich zeilen- oder spaltenweise. Für allgemeine Sortierungen nach mehreren Vektoren stellt Excel die Funktion SORTIEREN() zur Verfügung.

i Hinweis

Excels SORTIERENNACH()-Funktion kann einen Bereich zeilen- oder spaltenweise sortieren. Diese Funktion hat vier Parameter:

- Matrix der zu sortierende Bereich, der keine Matrix sein muss.
- Sortierindex die Spalten- oder Zeilennummer, nach der sortiert werden soll. Standardmässig wird die erste Spalte bzw. die erste Zeile angenommen.
- Sortierreihenfolge legt die Sortierreihenfolge fest. 1, um aufsteigend und -1, um absteigend zu sortieren.
- nach_Spalte Ein Wahrheitswert, ob die Spalten oder die Zeilen sortiert werden sollen. WAHR bedeutet, dass die Spalten (horizontal) sortiert werden sollen. FALSCH bedeutet, dass die Zeilen (vertikal) sortiert werden sollen. Standardmässig wird zeilenweise sortiert.

Die Funktion ermöglicht es, mehrere Vektoren auf einmal nach **mehreren** gemeinsamen Kriterien zu sortieren. Dazu müssen zuerst die Sortierkriterien identifiziert werden.

Die Funktion SORTIERENNACH() deckt den Spezialfall ab, wenn die Sortierung nach einem externen Kriterium erfolgen soll. Während SORTIEREN() erfordert, dass die Sortierindizes im zu sortierenden Bereich enthalten sind, können die Sortierindizes bei SORTIERENNACH() an einer beliebigen Stelle in der Arbeitsmappe liegen.

10.8.0.1. Schritt 1: Sortierkriterien festlegen.

Die Sortierkriterien sind durch die Werte im Sortierindex festgelegt, nach denen sortiert werden soll. Der Sortierindex ist ein Vektor mit einem Wert für eine Zeile bzw. Spalte der Sortiermatrix. Entlang der Werte im Sortierindex wird die Sortiermatrix sortiert.

In Excel können die Vektoren mit den Sortierkriterien an einer beliebigen Position in einer Arbeitsmappe liegen. Dabei müssen zwei Bedingungen erfüllt sein:

- 1. Der Sortierindex und die Sortiermatrix müssen die gleiche Länge haben.
- 2. Die Sortierindex und die Sortiermatrix müssen die gleiche Orientierung haben.

10.8.0.2. Schritt 2:

Im zweiten Schritt werden die zu sortierenden Vektoren ausgewählt.

In R wird dieser zweite Schritt automatisch auf die vorgegebene Stichprobe angewandt. In Excel können wir zusammenhängende Vektoren als "Matrix" an die SORTIERENNACH()-Funktion übergeben. Hängen die Vektoren nicht direkt zusammen, dann müssen mehrere Sortieroperationen mit den gleichen Referenzen auf die Sortierreferenzen durchgeführt werden.

In Excel wird die Sortierrichtung als Sortierreihenfolge bezeichnet und als separater Parameter für das jeweilige Sortierkriterium angegeben. Dabei steht 1 für die aufsteigende Sortierung und -1 für die absteigende Sortierung.

10.9. Rezepte

10.9.1. Fehlerwerte abfangen

Viele Excel-Funktionen geben einen Fehlerwert zurück, falls die Funktion kein gültiges Ergebnis ermitteln kann. Weil sich diese Fehlerwerte in Operationen fortpflanzen, müssen diese Werte durch einen geeigneten regulären Wert ersetzt werden. Das kann mit der folgenden Entscheidung erreicht werden.

```
= WENN(ISTFEHLER(A1); 0; A1)
```

Diese Operation ersetzt alle Fehlerwerte durch den Wert 0 und lässt alle anderen Werte unverändert.

Weil diese Entscheidung sehr oft vorkommt, gibt es die Funktion WENNFEHLER(), mit der die gleiche Operation einfacher ausgedrückt werden kann.

= WENNFEHLER(A1; 0)

10.9.2. Eine Zahl für genau eine Bedingung zurückgeben

Ein häufiger Spezialfall für Unterscheidungen ist die Auswahl von Zahlen, die genau einen logischen Ausdrück erfüllen. Solche Unterscheidungen geben im FALSCH-Fall 0 und im anderen Fall die gesuchte Zahl zurück. In diesem Spezialfall kann der Zielwert ohne Umweg über die WENN()-Funktion mit dem logischen Ausdruck multipliziert werden. Der logische Ausdruck liefert 1 für WAHR und 0 für FALSCH. Die Multiplikation mit 0 liefert immer 0. Die Multiplikation mit 1 liefert den Zielwert.

Das Beispiel gibt für die folgenden Werte alle Zahlen zurück, die grösser als 10 und kleiner als 20 sind.

Normalerweise würde diese Entscheidung durch die folgende Operation abgebildet:

= WENN((A1:A10 > 10) * (A1:A10 < 20); A1:A10; 0)

Warnung

Diese spezielle Fallunterscheidung sollte auf Korrektheit überprüft werden, wenn im WAHR-Fall der Wert 0 erlaubt ist. In diesem Fall wird der Wert 0 nicht vom logischen Ausdruck unterschieden.

Weil alle Werte Zahlen sind, handelt es sich um den Spezialfall, dass der WAHR-Wert eine Zahl und der FALSCH-Fall eine 0 ist. Für diesen Fall lässt sich die Formel vereinfachen, indem die gesuchten Werte mit dem logischen Ausdruck multipliziert werden:

```
= A1:A10 * (A1:A10 > 10) * (A1:A10 < 20)
```

Das Ergebnis beider Formeln sind die Werte {13;0;17;0;12;0;0;11;0;0}.

Damit dieses Rezept funktioniert, müssen alle Teile des logischen Ausdrucks genau die Werte FALSCH oder WAHR bzw. 0 oder 1 zurückgeben. Das ist notwendig, weil nur das neutrale Element die Zielwerte unverändert lässt. Eine direkte Übergabe von Zahlen im logischen Ausdruck verfälscht das Ergebnis, weil nicht mit dem neutralen Element gerechnet wird.

Soll für einen logischen Ausdruck nur der Wert 1 oder 0 zurückgegeben werden, dann kann der Rückgabebereich am Anfang der Formel weggelassen werden. Es wird dann nur der logische Ausdruck angegeben. Die Formel = (A1:A10 > 10) * (A1:A10 < 20) hat die Werte $\{1;0;1;0;1;0;0;1;0;0\}$ als Ergebnis.

10.9.3. Fehlerwerte vergleichen

Fehlerwerte können nicht direkt mit den Vergleichsoperatoren verglichen werden, weil Excel immer den ersten gefundenen Fehlerwert als Ergebnis einer Operation zurückgibt. Deshalb müssen Fehlerwerte zuerst in normale Werte konvertiert werden. Damit verschiedene Fehlerwerte miteinander verglichen werden können, müssen die verschiedenen Fehlerwerte zuerst in eindeutige Zahlen umgewandelt werden. Das übernimmt die Funktion FEHLER. TYP(). Diese Zahlen können anschliessend wie gewohnt weiter verarbeitet werden.

Das folgende Beispiel weist den gegebenen Fehlerwerten eine Fehlermeldung zu:

```
A
#NV
#WERT!
3
```

Die folgende Formel liefert die Werte {"Fehler: #NV"; "Fehlerhafter Wert", "Kein Fehler"}.

```
= WENNS(FEHLER.TYP(A1:A2) = 7; "Fehler: #NV";
            FEHLER.TYP(A1:A2) = 3; "Fehlerhafter Wert";
            WAHR; "Kein Fehler")
```

Weil alle Vergleiche die Gleichheit überprüfen, kann die Formel mit der Funktion ERSTERWERT() vereinfacht werden. Die Formel lautet dann:

Für diesen Schritt muss die Operation mit der Funktion WENNFEHLER() erweitert werden, weil die Funktion FEHLER.TYP() einen Fehler ausgibt, wenn der übergebene Wert kein Fehlerwert ist. Weil die Fehlertypen mit Werten grösser 0 durchnummeriert sind, bietet sich für reguläre Werte der Wert 0 an.

10.9.4. Filtern und Summen

Bis Juli 2020 mussten die Funktionen SUMMEWENN() oder SUMMEWENNS() verwendet werden, um Daten nach Kriterien zu summieren. Diese Funktionen haben allerdings den Nachteil, dass keine echten logischen Ausdrücke verwendet werden können. Seit Juli 2020 steht die FILTER()-Funktion zur Verfügung. Dadurch können echte logische Ausdrücke als Filterkriterien eingesetzt werden. Gleichzeitig hat sich die Bedeutung der Funktionen SUMMEWENN() und SUMMEWENNS() geändert (s. Kapitel 13).

Die folgende Formel sollten wir jetzt mit der Filter-Technik umschreiben.

Beispiel 10.9 (Summenwenn durch Filter-Summe ersetzen). Alt wurde geschrieben:

```
= SUMMEWENN(B2#; "< 0")
```

Neu ist die gleiche Funktion etwas ausführlicher:

```
= SUMME( FILTER(B2#; B2# < 0) )
```

Diese Schreibweise hat den Vorteil, dass die einzelnen Schritte nach dem Prinzip der Problemzerlegung getrennt werden und separat untersucht werden können. Das war in der alten Schreibweise nur indirekt möglich.

Das gleiche Prinzip gilt auch für ZÄHLENWENN() und ZÄHLENWENNS() anwenden. In diesem Fall wird entweder die Funktion ANZAHL() oder die Funktion ANZAHL2() als Aggregator eingesetzt.

Der grösste Vorteil ist aber, dass mit dieser Technik beliebige EXCEL-Aggregatoren mit gefilterten Daten eingesetzbar sind und nicht mehr auf die vordefinierten Aggregatoren eingeschränkt sind. Mit dem Filtern wird es ausserdem möglich, andere logische Ausdrücke als nur einen direkte Vergleich oder, im Fall von SUMMEWENNS() oder ZÄHLENWENNS(), mit Und verknüpfte Vergleiche durchzuführen.



Warnung

Wenn wir komplexe logische Ausdrücke mit Excels FILTER()-Funktion verwenden wollen, dann müssen wir für die logischen Operatoren die arithmetische Schreibweise für die logischen Ausdrücke verwenden!

11. Vektoroperationen



12. Matrix-Operationen



13. Kodieren und Gruppieren



14. Daten formen



Teil IV. Deskriptive Datenanalyse

15. Daten beschreiben



lack Work in Progress

16. Daten visualisieren



17. Pivot-Tabellen



Work in Progress

Pivot-Tabellen sind eine spezielle Form von Tabellen, die zur Datenanalyse verwendet werden. Pivot-Tabellen sind eine tabellarische Darstellung der Daten. Weil die Daten für Folgeberechnungen nicht zuverlässig zugegriffen werden können, sollten Pivot-Tabellen als Darstellungsform und nicht als Datenstruktur behandelt werden.

Anders als Wertetabellen (s. Definition 6.2) sind Pivot-Tabellen keine benannte Struktur und die dargestellten Werte können nicht über die Tabellenadressierung abgefragt werden. Es gibt zwar eine Reihe von Funktionen, die auf Pivot-Tabellen angewendet werden können, aber diese Funktionen lassen sich nicht konsistent mit anderen Funktionen der Datenanalyse verkettet werden.

Referenzen

- Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: A critical review of Net Promoter Score. Applied Stochastic Models in Business and Industry, 35(1), 138–151. https://doi.org/10.1002/asmb.2417
- Grisaffe, D. B. (2007). Questions about the ultimate question: conceptual considerations in evaluating Reichheld's Net Promoter Score (NPS). 20.
- Keiningham, T. L., Cooil, B., Andreassen, T. W., & Aksoy, L. (2007). A Longitudinal Examination of Net Promoter and Firm Revenue Growth. *Journal of Marketing*, 71(3), 39–51. https://doi.org/10.1509/jmkg.71.3.039
- Microsoft Support. (2023a). Excel specifications and limits. https://support.microsoft.com/engb/office/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3
- Microsoft Support. (2023b). Using structured references with Excel tables. https://support.microsoft.com/en-gb/office/using-structured-references-with-excel-tables-f5ed2452-2337-4f71-bed3-c8ae6d2b276e
- Reichheld, F. F. (2003). The One Number You Need to Grow. Harvard Business Review.