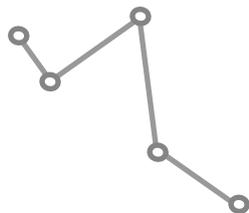
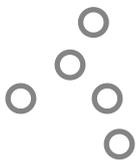


DATEN SÄUBERN IN GOOGLE SHEETS



DATEN SÄUBERN IN GOOGLE SHEETS

INHALTSVERZEICHNIS

1. Daten-Formate
2. Versteckte Infos
3. Conditional Formating
4. Daten einheitlich transformieren
5. Duplikate entfernen
6. Leere Zellen entfernen
7. Pivot Tabellen

Hat man einen Datensatz bekommen oder erstellt, ist der erste wichtige Schritt diesen für die Analyse zu säubern und zu strukturieren. Dies wird häufig vernachlässigt, ist aber ungemein wichtig, um verlässliche Ergebnisse zu erhalten. Daten säubern kann umständlich sein. Mit ein Tipps & Tricks können Daten jedoch viel effizienter bearbeitet werden. Das Lernmaterial listet die hilfreichsten Schritte in Google Sheets um Daten zu säubern. Erster Schritt: Erstelle immer eine Kopie der Originaldaten, bevor Du mit dem Säubern beginnst!

1. DATEN-FORMATE

Datensätze kommen in unterschiedlichen Formaten. Die häufigsten Formate, wenn man mit Google Sheets arbeitet, sind xls-files oder xlsx-files, die Standard Excel Formate, sowie CSV-Dateien. CSV steht für "comma separated values". Die Werte werden durch ein Komma oder manchmal auch ein Semikolon voneinander getrennt. JSON is ein weiterer Datentyp und steht für (JavaScript Object Notation).

Hilfreiche Quellen:

- Dieses Tutorial zeigt, wie sich JSON-Files in Goolge Sheets richtig importieren lassen.
<https://medium.com/@paulgambill/how-to-import-json-data-into-google-spreadsheets-in-less-than-5-minutes-a3fede1a014a>
- Mit dem Tool [konklone](https://konklone.io/json/) können JSON-Files online in CSV-Dateien konvertiert werden:
<https://konklone.io/json/>

2. VERSTECKTE INFOS

Erin Spreadsheet besteht aus Reihen und Spalten. Jedoch werden nicht immer alle ordentlich angezeigt. Wenn man mit einem neuen Datensatz arbeitet, sollte man zunächst überprüfen, ob es versteckte Spalten und Reihen enthält. Vielleicht wurden diese unbeabsichtigt mitveröffentlicht. Im Gegensatz zu Excel, hat Google Sheets allerdings keine Funktion, um alle Daten in einem Schritt anzeigen zu lassen.

Google Spreadsheet Columns



Zwei schwarze Pfeile zwischen zwei Spalten oder Reihen zeigen an, dass hier Informationen ausgeblendet werden. Wenn man auf die Pfeile klickt, werden die Spalten oder die Reihen wieder normal angezeigt.

3. CONDITIONAL FORMATING

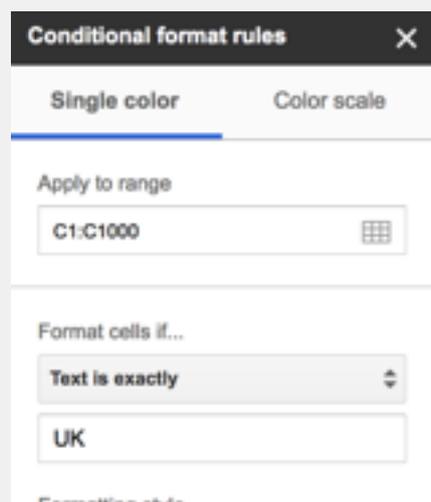
Zu überprüfen ob Datensätze einheitlich strukturiert sind, kann mühselig sein. Mit der „Conditional Formating“ Funktion lässt das Format von Spalten und der enthaltenen Werte schnell überprüfen. Angenommen Du hast einen riesigen Datensatz mit Informationen über Städte aus Großbritannien. Es ist klar, dass alle Städte im Vereinigten Königreich liegen, aber Du möchtest überprüfen, ob unterschiedliche Schreibweisen wie „United Kingdom“, „Great Britain“ oder „Britain“ verwendet wurden. Mit dem „Conditional Formating“ tool können unterschiedliche Bedingungen gesetzt werden. Falsch formatierte Werte werden so schnell identifiziert. Die Details der Funktionen werden in diesem youtube-video erklärt:

- https://www.youtube.com/watch?v=Cv_OEnNzdnU

Schritt 1:
Rechts-Klick auf eine Spalte



Schritt 2:
Die richtige Bedingung auswählen



Schritt 3:
Auf Inkonsistent prüfen



4. DATEN EINHEITLICH TRANSFORMIEREN

Google Sheets enthält einige hilfreiche Funktionen, um unstrukturierte Datensätze zu säubern. Sie helfen Werte einheitlich zu transformieren. Die wichtigsten Funktionen werden hier gelistet.

Function	Description
SÄUBERN()	Formatierungen wie <i>Kursiv</i> - oder Fettschrift werden entfernt
GLÄTTEN()	Alle Leerzeichen vor und hinter einem Wert werden entfernt
SPLIT()	Text wird nach einem festgelegten Trennzeichen getrennt. Jeder Textteil erhält eine Spalte.
GROSS()	Alle Buchstaben in einer Zelle werden in Großbuchstaben umgewandelt
KLEIN()	Alle Buchstaben in einer Zelle werden in Kleinbuchstaben umgewandelt
GROSS2()	Anfangsbuchstaben eines Wortes werden Groß geschrieben. Alle anderen in Kleinbuchstaben ausgegeben.

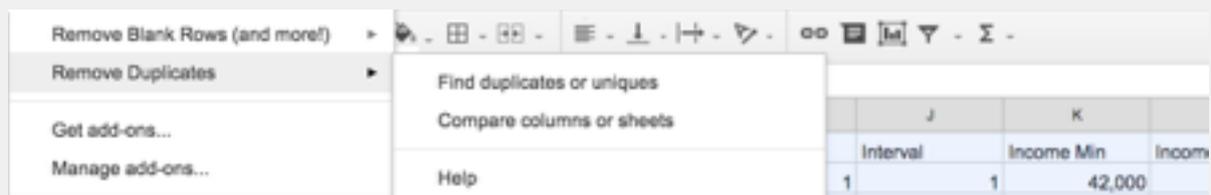
5. DUPLIKATE ENTFERNEN

Nicht immer aber häufig sollen Duplikate in einem Datensatz entfernt werden, da sie die Auswertung verzerren können. Es gibt unterschiedliche Möglichkeiten, um Duplikate in Google Sheets zu entfernen, unter Anderem ein Plugin, was genau für diese Aufgabe entwickelt wurde.

Duplikate entfernen:

Das Plugin wird über den Add-On Tab heruntergeladen und installiert.

1. Klicke anschließend erneut auf den Add-ons Tab
> Remove Duplicated > Find duplicates or uniques.



Wähle das "Remove Duplicates Plugin"

5. DUPLIKATE ENTFERNEN

2. Ein Pop-Up Fenster öffnet sich. In 4 Schritten kannst du die Details festlegen, wie Duplikate entfernt werden sollen. Zunächst kannst du den zu überprüfenden Bereich angeben.



Find duplicates or uniques

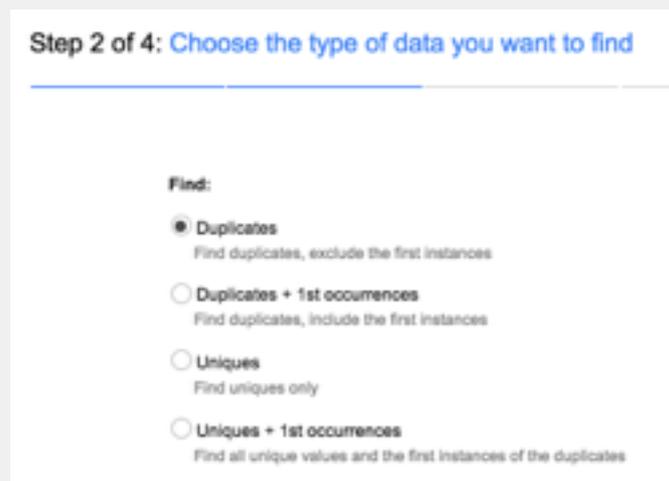
Step 1 of 4: **Select your sheet and range**

Select the range with your table:

A1:O61

Schritt 1: Wähle das Blatt & den Bereich aus

3. Als Nächstes muss angegeben werden, ob nach Duplikaten, nur der ersten Dublette, einzigartigen Werten oder dem ersten einzigartigen Wert in einem Spreadsheet gefiltert werden soll.



Step 2 of 4: **Choose the type of data you want to find**

Find:

- Duplicates**
Find duplicates, exclude the first instances
- Duplicates + 1st occurrences**
Find duplicates, include the first instances
- Uniques**
Find uniques only
- Uniques + 1st occurrences**
Find all unique values and the first instances of the duplicates

Schritt 2: Wähle die Daten nach denen gefiltert werden soll

5. DUPLIKATE ENTFERNEN

4. Im vierten Schritt müssen die zu überprüfenden Spalten festgelegt werden.

Step 3 of 4: Select columns to search in

Skip empty cells My table has headers
 Case sensitive

<input checked="" type="checkbox"/> Column	1st row content
<input checked="" type="checkbox"/> gender	male
<input checked="" type="checkbox"/> Party	CDU
<input checked="" type="checkbox"/> Organization	Hörzentrum Odenburg GmbH
<input checked="" type="checkbox"/> Job	Geschäftsführer

Schritt 3: Wähle die Spalte aus

5. Im letzten Schritt wird festgelegt, was mit den gefundenen Daten gemacht werden soll. Die Werte können farblich hervorgehoben, entfernt oder in eine neue Spalte kopiert werden.

Step 4 of 4: Choose what to do with the found values

Choose the action:

Fill with color

Add a status column

Copy to another location

Move to another location

Clear values

Delete rows within selection

Paste values to:

New sheet

New spreadsheet

Custom location

Schritt 4: Entscheide was mit den gefundenen Daten geschehen soll

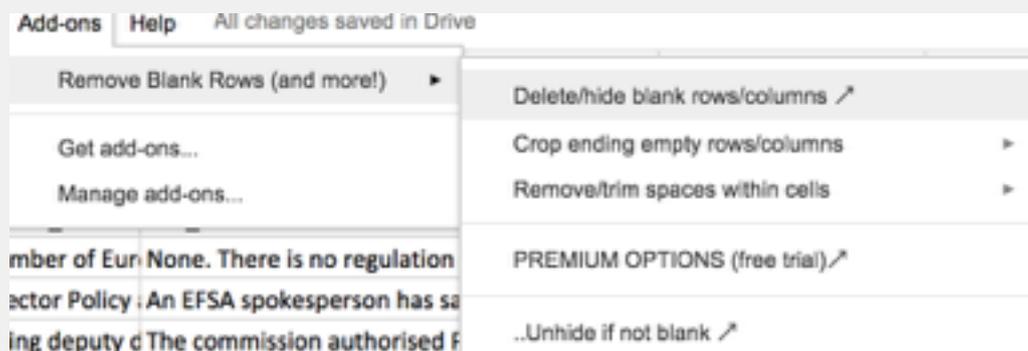
5. DUPLIKATE ENTFERNEN

Eine andere Methode, um Duplikate zu entfernen, bietet die `=unique()`-Funktion. Auf einen Bereich angewandt, gibt die Funktion nur einzigartige Werte zurück.

6. LEERE ZELLEN ENTFERNEN

Leere Reihen zu finden und per Hand zu löschen, kann sehr mühselig werden. Das Plugin „Remove blank rows“ ist hierfür hilfreich. Das Add-on wird wieder über „Add-ons > Get-Add ons“ eingebunden.

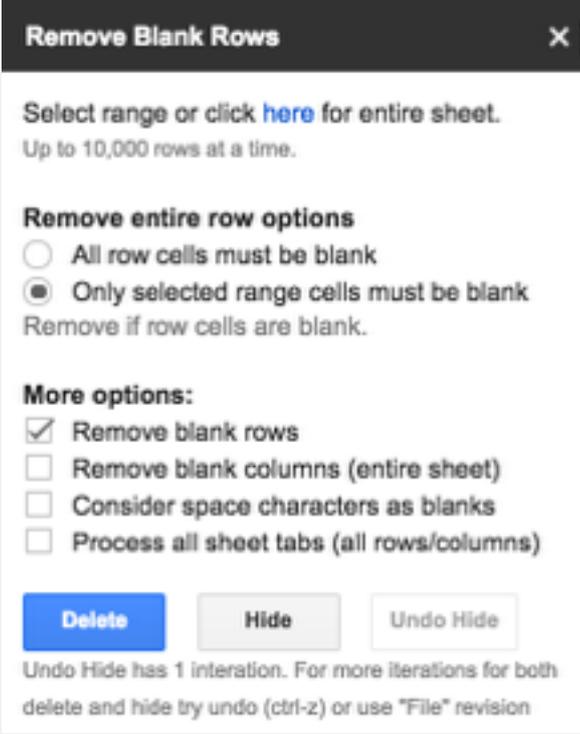
Nachdem das Plugin installiert ist klicke, auf Add-ons > Remove Blank Rows > Delete/hide blank rows/columns



Choose: Delete/hide rows/columns

6. LEERE ZELLEN ENTFERNEN

Als nächstes erscheint ein Kästchen auf der rechten Seite, indem festgelegt wird, was mit den identifizierten leeren Reihen geschehen soll. Hier können auch Lesezeichen als Leerzellen identifiziert und entfernt werden.



Remove Blank Rows ✕

Select range or click [here](#) for entire sheet.
Up to 10,000 rows at a time.

Remove entire row options

All row cells must be blank

Only selected range cells must be blank
Remove if row cells are blank.

More options:

Remove blank rows

Remove blank columns (entire sheet)

Consider space characters as blanks

Process all sheet tabs (all rows/columns)

Delete **Hide** **Undo Hide**

Undo Hide has 1 iteration. For more iterations for both delete and hide try undo (ctrl-z) or use "File" revision

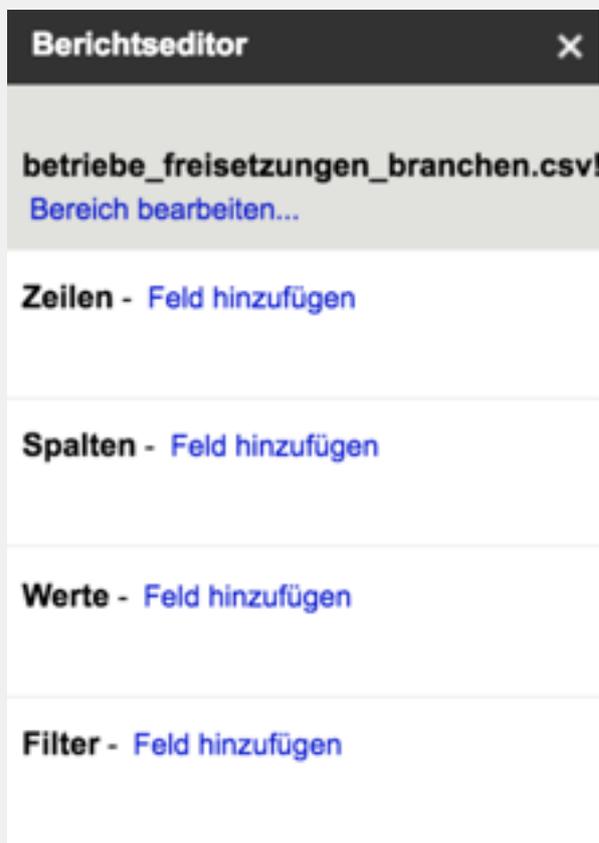
Auswahl: Was soll als Leerzeile identifiziert werden

7. PIVOT TABELLEN

Mittels der Pivot-Tabelle lassen sich Spalten aus einem Datenset zusammenfassen und miteinander in Verbindung setzen. Dadurch entstehen Kreuztabellen und man erhält einen schnellen Überblick, was in einem Datenset enthalten ist.

1. Der Bereichsfilter:

Der Editor



Die Optionen

Im Bereichseditor kann die Zusammensetzung der Pivot-Tabelle geändert werden.

Feld hinzufügen: Spalten aus der Datengrundlage können über den Button in die Pivot-Tabelle eingefügt werden

Zeilen: Die eingefügten Daten werden als Zeilen, untereinander dargestellt.

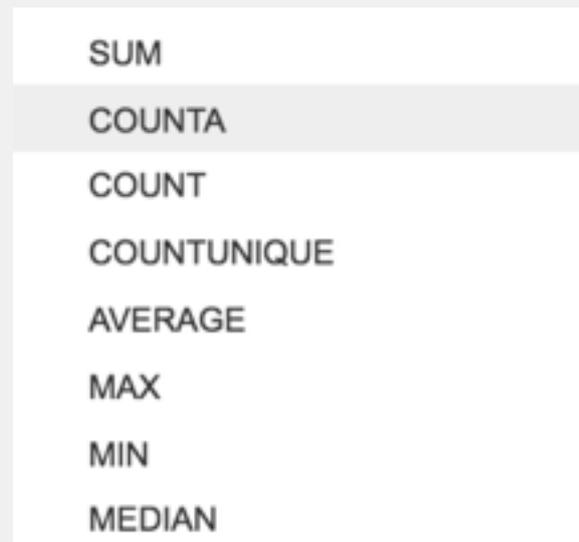
Spalten: Die eingefügten Daten werden als horizontal nebeneinander angeordnet

Werte: Die eingefügte Spalte wird hier als Ergebnis für die zuvor ausgewählten Spalten angezeigt.

Filter: Über den Filter können bestimmte Werte aus einer Spalte aus der Analyse ausgeschlossen werden.

7. PIVOT TABELLEN

2. Zusammenfassung der Werte:



Die Werte die in einer Pivot-Tabelle analysiert werden können auf unterschiedliche Weise *zusammengefasst* werden. Klickt man auf das dreieckige Pfeil werden mehrere Funktionen zur Zusammenfassung aufgelistet.

SUM: Werte werden summiert. (Nur möglich bei Zahlenwerten bestehen)

COUNTA: Textwerte können mit COUNTA zusammengefasst werden

COUNTUNIQUE: Fasst nur eindeutige Werte zusammen

AVERAGE: Gibt den Durchschnitt wieder

MIN/MAX: Geben den niedrigsten/höchsten Wert wieder

MEDIAN: Gibt den Mittelwert wieder



Die Datenschule vermittelt gemeinnützigen Organisationen die nötigen Fähigkeiten, Daten und Technologien zu verstehen, um sie zielgerichtet für ihre gesellschaftlichen Aufgaben einzusetzen.



OPEN
KNOWLEDGE
FOUNDATION
DEUTSCHLAND

Die Open Knowledge Foundation Deutschland ist ein gemeinnütziger Verein, der sich für offenes Wissen, offene Daten, Transparenz und Beteiligung einsetzt.

Mehr zu unseren Projekten &
Workshops: datenschule.de
E-Mail: info@datenschule.de |
Telefon: 030-57703666-2