

Datenexport und Vorbereitung der Datenanalyse

Diese Anleitung befasst sich mit der Datenvorbereitung für die Darstellung im Diagramm. Sie erfahren, wie der Zugriff auf die Datenbanken und wie der Datenexport von den Datenbanken funktioniert. Weiterhin erhalten Sie eine Anleitung, welche Schritte zur Vorbereitung für die Datenanalyse notwendig sind. Dafür werden zwei Programme benötigt. Zum einen ist eine Version von Excel notwendig, um die Vorbereitungen für die Datenanalyse zu treffen und anschließend die erstellten Tabellen als Diagramme zu präsentieren. Diese Handreichung orientiert sich an der Version Microsoft Office Excel 2016. Bei anderen Versionen von Excel können die folgenden, erklärten Schritte abweichen. Zum anderen benötigen Sie die kostenlose Software „DB Browser for SQLite“. Sie ermöglicht die Öffnung und den Datenexport von den Datenbanken. Diese Software kann unter folgendem Link kostenfrei heruntergeladen werden: <https://sqlitebrowser.org/dl/>.

1 Zugriff auf die Datenbanken

Die MicroSD-Karte des Raspberry Pi hat eine Größe von 8 GB. Das bedeutet, die MicroSD-Karte kann für eine lange Zeit Daten speichern. Je nachdem, was für Zeiträume bei der Datenanalyse visualisiert werden sollen, können die Daten in selbst festgelegten Abständen (wöchentlich, monatlich) heruntergeladen und gesichert werden. Sie haben eine manuelle Möglichkeit, auf die Datenbanken der Umweltmessstation zuzugreifen. Dieser Weg wird im Folgenden erläutert.

Der manuelle Weg: Hierbei wird die MicroSD-Karte vom Raspberry Pi entnommen und mithilfe eines Adapters an den PC/Laptop angeschlossen. In dem Unterordner *database* befinden sich die beiden Datenbanken **sht31** und **Rain**, die dann auf den eigenen Computer kopiert werden können. Mithilfe des Programms „DB Browser for SQLite“ können die zwei Datenbanken geöffnet und deren Daten beispielsweise als CSV-Datei exportiert werden. Dies wird genauer im nächsten Kapitel dieser Handreichung erklärt.

2 Export der Daten

Schritt 1: Datenbank öffnen

Starten Sie das Programm „DB Browser for SQLite“. Klicken Sie anschließend auf: *Datei* → *Datenbank öffnen* → Datenbank (*Rain* / *sht31*) auswählen.

Schritt 2: Datenbank als CSV-Datei exportieren

Damit die Datenbanken in Excel bearbeitet werden können, müssen sie zuerst als CSV-Dateien exportiert werden. Klicken Sie bitte dazu auf: *Datei* → *Export* → *Tabelle(n) als CSV-Datei*. Dabei werden sie aufgefordert, verschiedene Einstellungen anzugeben. Die Abbildung 1 stellt die gewünschten Exporteinstellungen dar.



Abbildung 1 Einstellungen für den Export als CSV-Datei

Schritt 3: CSV-Datei in Excel öffnen

Um mit den gespeicherten CSV-Dateien arbeiten zu können, müssen diese in einer Excel-Arbeitsmappe in zwei verschiedenen Blättern geöffnet werden. Klicken Sie im jeweiligen Blatt auf: *Daten* → *aus Text/CSV* → CSV-Datei auswählen → *laden*.

Ein Problem beim Importieren der Daten tritt in der Spalte *value* auf. Diese Werte werden ohne die Dezimaltrennung übernommen, sodass aus „24.1“ „241“ wird. Dies kann behoben werden, indem eine neue Tabellenspalte, z. B. *value_2*, angelegt wird. Dort werden die Werte durch 10 dividiert. Die Werte der Datenbank *Rain* müssen noch mit 0,2 multipliziert werden, um die insgesamten Niederschlagsmenge in mm zu berechnen. Die folgende Tabelle zeigt die Berechnungsvorschrift für die Spalte *value_2* und für die jeweilige Datenbank:

<u>sht31</u>	<u>Rain</u>
=Zelle <i>value</i> /10	=(Zelle <i>value</i> /10)*0,2

	A	B	C	D	E
1	type	value	time	uploaded	value_2
2	T	241	13.06.2019 11:26	0	=B2/10
3	H	527	13.06.2019 11:26	0	

Abbildung 2 Anlegen der Spalte *value_2*

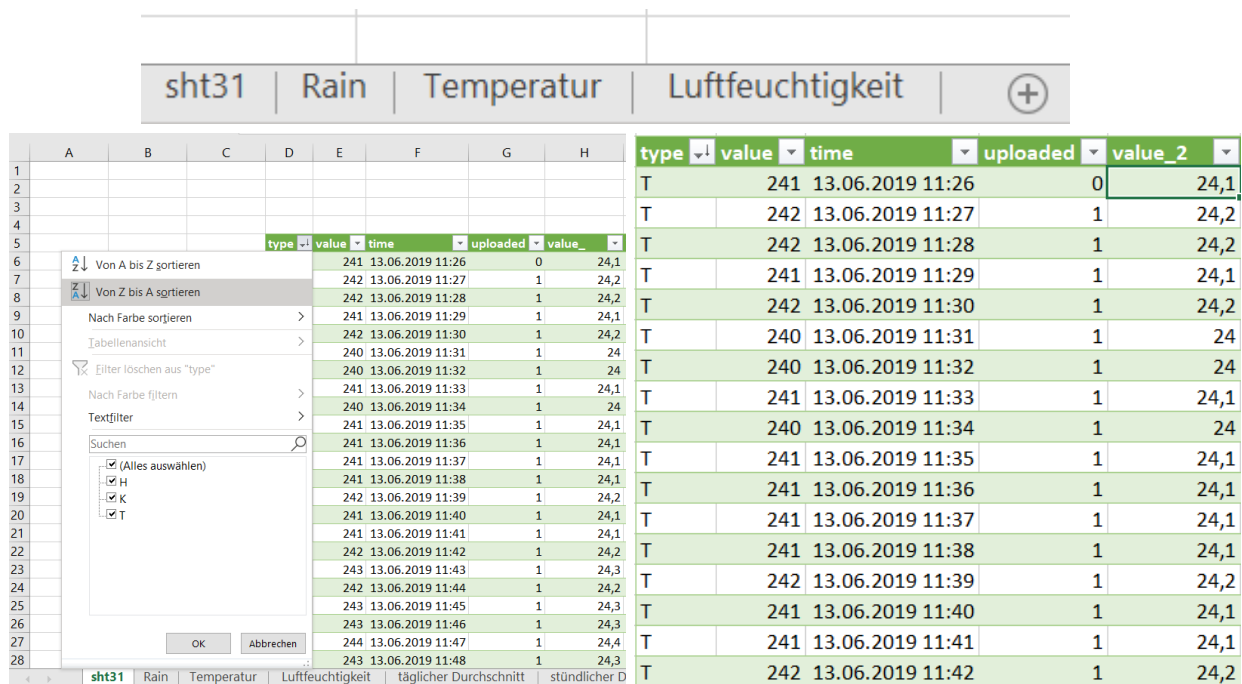
Gegebenenfalls wird die Berechnungsvorschrift/ Funktion nicht für jede Zelle übernommen. Wenn das der Fall ist, klicken Sie bitte in die Zelle und klicken Sie doppelt auf das kleine schwarze Kreuz in der rechten, unteren Ecke. Die Funktion wird nun bis zum Tabellenende übernommen.

Schritt 4: Anlegen neuer Blätter für die gemessenen Umweltfaktoren Temperatur und Luftfeuchtigkeit

Dieser Schritt entfällt für die Datenbank *Rain*.

Für die Datenanalyse ist es von Vorteil, wenn die gemessenen Umweltfaktoren in unterschiedlichen Blättern aufbereitet werden. Das heißt also, zu den zwei Blättern für die zwei Datenbanken erstellen Sie zwei neue für die Temperatur und die Luftfeuchtigkeit. Klicken Sie dafür unten links in Excel auf das Plus-Zeichen neben den beiden Datenbanken *sht31* und *Rain*.

Für das Kopieren der Daten bietet es sich an, die Spalte *type* der Datenbank *sht31* von „A bis Z sortieren“ zu lassen. Im Anschluss können dann die jeweiligen Werte kopiert und in das dazugehörige Blatt eingefügt werden. Für die Markierung der Werte bietet die Shift-Taste eine große Vereinfachung. Mithilfe dieser Taste werden alle Zeilen zwischen der ersten und letzten angeklickten Zeile markiert.



type	value	time	uploaded	value_2
T	241	13.06.2019 11:26	0	24,1
T	242	13.06.2019 11:27	1	24,2
T	242	13.06.2019 11:28	1	24,2
T	241	13.06.2019 11:29	1	24,1
T	242	13.06.2019 11:30	1	24,2
T	240	13.06.2019 11:31	1	24
T	240	13.06.2019 11:32	1	24
T	241	13.06.2019 11:33	1	24,1
T	240	13.06.2019 11:34	1	24
T	241	13.06.2019 11:35	1	24,1
T	241	13.06.2019 11:36	1	24,1
T	241	13.06.2019 11:37	1	24,1
T	241	13.06.2019 11:38	1	24,1
T	242	13.06.2019 11:39	1	24,2
T	241	13.06.2019 11:40	1	24,1
T	241	13.06.2019 11:41	1	24,1
T	242	13.06.2019 11:42	1	24,2
T	243	13.06.2019 11:43	1	24,3
T	242	13.06.2019 11:44	1	24,2
T	243	13.06.2019 11:45	1	24,3
T	243	13.06.2019 11:46	1	24,3
T	244	13.06.2019 11:47	1	24,4
T	243	13.06.2019 11:48	1	24,3
T	242	13.06.2019 11:42	1	24,2

Abbildung 3 Anlegen neuer Blätter für Umweltfaktoren

3 Vorbereitung für die Datenanalyse und Darstellung

Aufgrund der minütlichen Messung der Umweltmessstation sind die Datenbanken sehr hochauflösend. Excel stellt viele Funktionen bereit, um das Arbeiten mit großen Datensätzen zu erleichtern. Vor der Visualisierung muss entschieden werden, was dargestellt werden soll. In dem Fall der Umweltmessstation gibt es zwei Optionen, die Daten zielführend darzustellen. Das ist zum einen eine detaillierte Tagesverlaufskurve mit stündlichen Durchschnittswerten der Temperatur und Luftfeuchtigkeit sowie der stündlichen Niederschlagssumme und zum anderen eine Darstellung von täglichen Durchschnittswerten und Niederschlagssummen über einen längeren Zeitraum hinweg.

Die folgenden Schritte zeigen eine Bearbeitung der Tabellen, um die Durchschnitts- und Summenwerte zu berechnen. Dabei gelten sowohl für die Tagesverlaufskurve als auch die Zeitreihe dieselben Schritte. Für die Berechnung werden verschiedene Vergleichskriterien benötigt: Tag, Monat und Jahr. Nur für die Tagesverlaufskurve gibt es einen Vergleichswert mehr: Stunde. Es wird im Folgenden von einer Originaltabelle und einer neuen Tabelle in einem neuen Blatt gesprochen, die Originaltabelle ist grün, die neue Tabelle ist blau.

3.1 Durchschnittswert und Summenwert

Funktionen: MITTELWERTWENNS, SUMMEWENNS, TAG, MONAT, JAHR, STUNDE, ISTFEHLER

Schritt 1: Anlegen neuer Spalten für Tag, Monat, Jahr und Stunde in Originaltabelle

Damit der Mittelwert der Temperatur- und Luftfeuchtigkeitsdaten einfach bestimmt werden kann, kann die Funktion MITTELWERTWENNS von Excel genutzt werden. Diese berechnet den Mittelwert, wenn verschiedene Kriterien erfüllt werden. Für diesen Fall werden die Vergleichskriterien über das Datum und die Uhrzeit entwickelt. Dazu müssen in der Originaltabelle von *sht31* und *Rain* vier neue Spalten angelegt werden, die jeweils für den Tag, den Monat, das Jahr und die Stunde stehen. Mithilfe der Funktionen TAG, MONAT, JAHR und STUNDE können die Zahlen aus der Bezugzelle ermittelt werden.

[illegible]

Abbildung 4 Anlegen von neuen Spalten in Originaltabelle

Schritt 2: Anlegen eines neuen Blattes mit neuer Tabelle

Nun werden zwei neue Blätter geöffnet, die beispielsweise mit „täglicher Durchschnitt“ und „stündlicher Durchschnitt“ benannt werden. Die folgenden Abbildungen 5 und 6 zeigen, wie diese Blätter mit den neuen Tabellen aufgebaut werden sollen. Bei den stündlichen Durchschnitten ist es wichtig, dass neben der Datumsspalte noch eine weitere Uhrzeitspalte angelegt wird, da diese für die Diagrammerstellung notwendig ist. Die Zellen der Uhrzeitspalte müssen als Uhrzeit formatiert werden. Dazu markieren Sie die Zellen und klicken: Rechtsklick → *Zellen formatieren* → *Uhrzeit* → *Typ: 13:30* → *ok*.

Datum	Mittelwert Temp	Mittelwert Luft	Niederschlagssumme	Tag	Monat	Jahr
13.06.2019						
14.06.2019						
15.06.2019						
16.06.2019						
17.06.2019						
18.06.2019						
19.06.2019						
20.06.2019						
21.06.2019						
22.06.2019						
23.06.2019						

Abbildung 5 Matrix für tägliche Durchschnitts- und Summenwerte

Datum	Uhrzeit	Mittelwert Temp [°C]	Mittelwert Luft [%]	stündliche Niederschlagssumme [mm]	Tag	Monat	Jahr	Stunde
13.06.2019 00:00	0:00							
13.06.2019 01:00	1:00							
13.06.2019 02:00	2:00							
13.06.2019 03:00	3:00							
13.06.2019 04:00	4:00							
13.06.2019 05:00	5:00							
13.06.2019 06:00	6:00							
13.06.2019 07:00	7:00							
13.06.2019 08:00	8:00							
13.06.2019 09:00	9:00							

Abbildung 6 Matrix für stündliche Durchschnitts- und Summenwerte

Schritt 3: Funktion MITTELWERTWENNS anwenden

Jetzt werden die Mittelwerte der Temperatur und der Luftfeuchtigkeit berechnet. Dafür klicken Sie zuerst in die 1. Zelle der neuen Tabelle, in der der Mittelwert des 1. Tages stehen soll und klicken anschließend auf f_x vor der Excel-Bearbeitungszeile. Es öffnet sich ein Fenster, in welchem die Funktionswerte eingegeben werden müssen (vgl. Abb. 7). Beim Mittelwert_Bereich wird die Spalte *value_2* aus der Originaltabelle übergeben. Beim Kriterien_Bereich1 wird die Spalte *Tag* aus der Originaltabelle übergeben, die als erstes Vergleichskriterium dient. Bei Kriterien1 wird die Zelle des 1. Tages angegeben (Spalte *Tag* in der neuen Tabelle). Der Kriterien_Bereich2 bzw. Kriterien2 stellen den Monat und Kriterien_Bereich3 bzw. Kriterien3 das Jahr dar. Kriterien_Bereich4 bzw. Kriterien4 für die Stunde wird nur bei den stündlichen Durchschnitts- und Summenwerten benötigt. Halten Sie die Werte bei der Übergabe für die „_Bereiche“ per Hand mit dem \$-Zeichen fest (vgl. Abb. 7). Klicken Sie zum Übergeben der Werte jeweils auf den Pfeil am Ende der Zeile und markieren Sie den gewünschten Wertebereich.

Funktionsargumente

MITTELWERTWENNS

Mittelwert_Bereich: Temperatur!\$I\$11:\$J\$10717
 Kriterien_Bereich1: Temperatur!\$I\$11:\$J\$10717
 Kriterien1: H9
 Kriterien_Bereich2: Temperatur!\$K\$11:\$K\$10717
 Kriterien2: I9
 = 23,14509284
 Sucht den Mittelwert (arithmetisches Mittel) für die Zellen, die durch bestimmte Bedingungen oder Kriterien angegeben sind.
 Mittelwert_Bereich: sind die tatsächlichen Zellen, die zum Suchen des Mittelwerts verwendet werden sollen.
 Formelergebnis = 23,14509284
 Hilfe für diese Funktion
 OK
 Abbrechen

Funktionsargumente

MITTELWERTWENNS

Mittelwert_Bereich: Luftfeuchtigkeit!\$H\$13:\$H\$10719
 Kriterien_Bereich1: Luftfeuchtigkeit!\$I\$13:\$J\$10719
 Kriterien1: H9
 Kriterien_Bereich2: Luftfeuchtigkeit!\$I\$13:\$J\$10719
 Kriterien2: I9
 = 58,87798408
 Sucht den Mittelwert (arithmetisches Mittel) für die Zellen, die durch bestimmte Bedingungen oder Kriterien angegeben sind.
 Mittelwert_Bereich: sind die tatsächlichen Zellen, die zum Suchen des Mittelwerts verwendet werden sollen.
 Formelergebnis = 58,87798408
 Hilfe für diese Funktion
 OK
 Abbrechen

Abbildung 7 Funktionsargumente MITTELWERTWENNS

Schritt 4: Funktion SUMMEWENNS anwenden

Für SUMMEWENNS gilt dasselbe wie für MITTELWERTWENNS. Hier wird jedoch die Spalte *value_2* aus der Datenbank *Rain* für den Summe_Bereich übergeben. Die Abbildung 8 zeigt das Fenster für die Funktionsargumente.

Abbildung 8 Fenster für Funktionsargumente SUMMEWENNS

Schritt 5: Zellen formatieren

Die Durchschnitts- und Summenwerte werden ggf. als Zahlen angezeigt, die mehr Nachkommastellen zeigen, als erwünscht ist. Dies kann geändert werden, indem die Zellen formatiert werden. Dazu klicken Sie bei den markierten Zeilen: Rechtsklick → *Zellen formatieren* → *Zahl* → *Dezimalstellen: 1* → *ok*.

Schritt 6: Fehlerspalte einfügen

Es kann beim Import der Daten aus der Datenbank Tage bzw. Stunden geben, an denen keine Werte gemessen wurden. Für diese Tage wird eine weitere Spalte erstellt, die die Funktion ISTFEHLER enthalten soll. Mithilfe dieser Funktion können die Tage bzw. Stunden angezeigt werden, die Werte besitzen. Klicken Sie dafür in die 1. Zeile der neuen Spalte und anschließend auf *f_x*. Wählen Sie ISTFEHLER aus. In dem Fenster für die Funktionsargumente übergeben Sie den Mittelwert für die Temperatur aus der 1. Zeile. Wenn es keinen Mittelwert für die Temperatur gibt, gibt es auch keinen Mittelwert für die Luftfeuchtigkeit. Diese Funktion wird nun automatisch für die gesamte Tabelle übernommen. Wenn in der Spalte ein WAHR steht, wurden keine Werte gemessen. Steht dort ein FALSCH, sind Werte für den Tag bzw. die Stunde vorhanden.

Abbildung 9 Funktion ISTFEHLER

Steht dort ein FALSCH, sind Werte für den Tag bzw. die Stunde vorhanden. Mit dem automatischen Filter können Sie nun die Tage bzw. Stunden ausblenden, die keine Messwerte besitzen.