



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 \rightarrow Datenbank öffnen \rightarrow 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 \rightarrow Export \rightarrow Tabelle(n) als CSV-Datei. Dabei werden sie aufgefordert, verschiedene Einstellungen anzugeben. Die Abbildung 1 stellt die gewünschten Exporteinstellungen dar.

Daten als CSV exportieren			?	\times
Tabelle(n)				
Spaltennamen in der ersten Zeile	\checkmark			
Feld-Separator	;	\sim		
String-Zeichen		\sim		
Zeilenumbruchs-Zeichen	Windows: CR+LF ($r\n)$	\sim		
	Speich	ern	Abbrech	nen

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 \rightarrow aus Text/CSV \rightarrow CSV-Datei auswählen \rightarrow *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 insgesamte Niederschlagsmenge in mm zu berechnen. Die folgende Tabelle zeigt die Berechnungsvorschrift für die Spalte value 2 und für die jeweilige Datenbank:

<u>sht31</u>						Rain				
=Zelle value/10					=(Zel	le value/10)	*0,2			
		Α	В	С		D	E			
	1	type 💌	value 💌	time	-	uploaded	value_2	-		
	2	Т	241	13.06.2019	11:26		0 = <mark>B2/1</mark> 0		-	
	3	н	527	13.06.2019	11:26		0		-	
	∆hhi	lduna 2	Anlegen de	r Snalte 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.

			sht	31		Rain	Te	mper	atur		Luft	feuch	tigkeit		(+)	
	A	В		С	D	E	F	G	Н	typ	e 🖵 v	/alue 💌	time	*	uploaded 🔽	value_2 💌
1 2										Т		241	13.06.2019 1	1:26	C	24,1
3										Т		242	13.06.2019 1	1:27	1	24,2
5	_				type 🚽	value 💌	time	uploaded 💌	value_ 💌	т		242	13.06.2019 1	1:28	1	24,2
6	₽↓	Von A bis Z s	ortieren			241	13.06.2019 11:2	6 0 7 1	24,1	т		241	13.06.2019 1	1:29	1	24.1
8	Z↓	Von Z bis A s	ortieren			242	13.06.2019 11:2	8 1	24,2	т		242	13 06 2019 1	1.30	1	24.2
9	_	Nach Farbe sor	tieren		>	241	13.06.2019 11:2	9 1	24,1	-		272	13.00.2013 1	1.50		27,2
10	-	Tabellenansich	t		>	242	13.06.2019 11:3	0 1	24,2	1		240	13.06.20191	1:31	1	. 24
12	- Tx	Eilter löschen	aus "type			240	13.06.2019 11:3	2 1	24	Т		240	13.06.2019 1	1:32	1	24
13		Nach Farbe filt	ern		>	241	13.06.2019 11:3	3 1	24,1	т		241	13 06 2019 1	1.33	1	24.1
14	_	Textfilter			>	240	13.06.2019 11:34	4 1	24	-		271	15.00.2015	1.55		27,1
15	_	Cushen			0	241	13.06.2019 11:3	5 1 6 1	24,1	Т		240	13.06.2019 1	1:34	1	. 24
17		Allos au	cwählon)		~	241	13.06.2019 11:3	7 1	24,1	Т		241	13.06.2019 1	1:35	1	24,1
18		- I H	swamen)			241	13.06.2019 11:3	8 1	24,1	т		244	12.00.2010.1	1.20	4	24.1
19		🗹 K				242	13.06.2019 11:3	9 1	24,2	1		241	13.06.2019 1	1:30	1	. 24,1
20	_					241	13.06.2019 11:4	0 1	24,1	T		241	13.06.2019 1	1:37	1	24,1
21	_					241	13.06.2019 11:4	1 1	24,1	т		2/11	12 06 2010 1	1.20	1	2/1
22	_					242	13.06.2019 11:4	3 1	24,2			241	13.00.20191	1.50	1	. 24,1
24						242	13.06.2019 11:4	4 1	24,2	T		242	13.06.2019 1	1:39	1	24,2
25						243	13.06.2019 11:4	5 1	24,3	т		241	13.06.2019 1	1:40	1	24.1
26	_			_		243	13.06.2019 11:4	6 1	24,3	-			10.00.2019		-	,_
27	_		OK	A	obrechen	244	13.06.2019 11:4	7 1	24,4	1		241	13.06.20191	1:41	1	24,1
20	•	sht31 Raii	n Ten	nperatur	Luftf	euchtigkei	it täglicher D	8 1 urchschnitt	stündlicher D	Т		242	13.06.2019 1	1:42	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 Bezugszelle ermittelt werden.

А	В	С		D	E	F	G	
	type	value	time		uploaded	value 2		
	Н	5	27 13.0	6.2019 11:26	0	52,7	=TAG(D5)	
	Н	5	533 13.0	6.2019 11:27	1	53,3		
type	🕂 value 💌	time 💌	uploaded 💌	value_2 🔽 T	ag 🔽 Mor	nat 💌 Jahr	🔹 Stunde	•
Т	241	13.06.2019 11:26	0	24,1	13	6	2019	11
Т	242	13.06.2019 11:27	1	24,2	13	6	2019	11
Т	242	13.06.2019 11:28	1	24,2	13	6	2019	11
т	241	13.06.2019 11:29	1	24,1	13	6	2019	11
Т	242	13.06.2019 11:30	1	24,2	13	6	2019	11
Т	240	13.06.2019 11:31	1	24	13	6	2019	11
т	240	13.06.2019 11:32	1	24	13	6	2019	11
Т	241	13.06.2019 11:33	1	24,1	13	6	2019	11
Т	240	13.06.2019 11:34	1	24	13	6	2019	11
Т	241	13.06.2019 11:35	1	24,1	13	6	2019	11
Т	241	13.06.2019 11:36	1	24,1	13	6	2019	11
Т	241	13.06.2019 11:37	1	24,1	13	6	2019	11
т	241	13.06.2019 11:38	1	24.1	13	6	2019	11

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 \rightarrow Zellen formatieren \rightarrow Uhrzeit \rightarrow Typ: 13:30 \rightarrow ok.

Datum 💽	Mittelwert Temp	 Mittelwert Luft 	Niederschlagssumme 💌	Tag 💌	Monat 🛛 💌	Jahr	Ŧ
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.



Abbildung 7 Funktionsargumente MITTELWERTWENNS

TU Dresden | Juniorprofessur für Didaktik der Geographie und Umweltkommunikation | <u>sensor@tu-dresden.de</u> © <u>CC BY-SA 3.0</u> Projekt "SENSOr – Smart Energy Smart Schools", "SmartRain_02_Datenexport_und_Vorbereitung_der_Datenanalyse" Dieses Projekt wird gefördert durch: Deutsche Bundesstiftung Umwelt (DBU), Osnabrück





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.

Funktionsargumente				?	\times
SUMMEWENNS					
Summe_Bereich	Rain!\$G\$7:\$G\$11302	1	=	{0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;	^
Kriterien_Bereich1	Rain!\$H\$7:\$H\$11302	1	=	{13;13;13;13;13;13;13;13;13;13;13;13;13;1	
Kriterien1	Н9	1	=	13	
Kriterien_Bereich2	Rain!\$I\$7:\$I\$11302	1	=	{6;6;6;6;6;6;6;6;6;6;6;6;6;6;6;6;6;6;6;	
Kriterien2	19	1	=	6	\sim
Addiert die Zellen, die von eine S	r bestimmten Gruppe von Bedingu umme_Bereich: sind die tatsächl	ngen oder K ich zu addie	= (riter renc	1,8 ien angegeben sind. len Zellen.	
Formelergebnis = 1,8					
Hilfe für diese Funktion				OK Abbreck	nen

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 \rightarrow Zellen formatieren \rightarrow Zahl \rightarrow Dezimalstellen: $1 \rightarrow ok$.

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 ISTFEH-LER 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 fx. 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

unkti	onsa	rgumente						? >
ISTFE	HLER							
			Wert E	9		1 = 23,145	09284	
						= FALSC	н	
rüft, ol	b ein	Wert ein Fe	ehler ist, und gib	t WAHR oder F	ALSCH zurück.	171250		
			-	10/	- West day and "ft.			
				Nam	er wert, der gepruπ w e. der sich auf eine Ze	elle. Formel oder eine	kann eine Zeile sein, eine n Wert bezieht.	Formel oder
					.,			
rmel	eraeh	nis = FAI	SCH					
men	ergeb	1113 - TAL	.5011					
lfe fü	ir dies	e Funktion					ОК	Abbrechen
						410		
			-					
	×	√ fx	=ISTFEHLER(E9)					
	В	C	D	E	F	G H	I I J	K
			Datum • Mitt	telwert lemp 23.1	Mittelwert Luft Niede	rschlagssumme 💌 lag	2 Nach Größe sortieren (aufstei	gend)
			14.06.2019	23,1	63.0	1,0	Z↓ Nach Größe sortieren (abstei	end)
			15.06.2019	#DIV/01	#DIV/01	0	Mark Fasher and Source	
			16.06.2019	#DIV/01	#DIV/0!	0	Nach Farbe sortjieren	
			17.06.2019	#DIV/01	#DIV/01	0	[abellenansicht	>
			18.06.2019	#DIV/01	#DIV/01	0	Filter löschen aus "Filterspalte	e
			19.06.2019	#DIV/01	#DIV/01	0	IX 2000 Construction of the	
			20.06.2019	#DIV/01	#DIV/0!	0	Nach Farbe filtern	>
			21.06.2019	23,7	56,8	1,2	Zahlenfilter	>
			22.06.2019	22,7	56,5	0,6	Suchen	0
			23.06.2019	22,8	53,9	0,8	Suchen	~
			24.06.2019	23,4	48,2	2,2	 (Alles auswählen) 	
			25.06.2019	24,5	58,7	4	I FALSCH	
			26.06.2019	25,7	63,4	4,2	WALIK	
			27.06.2019	24,8	47,8	3,2		
			28.06.2019	#DIV/01	#DIV/01	0		
			29.06.2019	#DIV/01	#DIV/01	0		
			30.06.2019	#DIV/0!	#DIV/0!	0		
			01.07.2019	#DIV/0!	#DIV/0!	0		
						-		
			02.07.2019	#DIV/01	#DIV/0	0		

Abbildung 9 Funktion ISTFEHLER

steht, wurden keine Werte gemessen. 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.