

AirQ Sensor Data Analytics



Dokumentation

Inhaltsverzeichnis

1.	Hintergrund und Ziel dieses Projekts.....	2
2.	Verwendete Sensorik.....	2
3.	Technische Umsetzung	3
4.	Projektablauf	4
4.1	Informationen einholen.....	4
4.2	Spezifikation der Schnittstelle	4
4.3	Entwicklung der Schnittstelle	5
4.4	Schnittstelle Überarbeiten.....	5
4.5	Korrelationsanalyse Labordaten	5
4.6	Erweiterung Funktionalität Programm.....	5
4.7	Auswertung der Daten	5
4.8	Installation Laborcomputer	6
5.	Analyseergebnisse der Daten	7
5.1	Analyse der Messdaten des Labors	8
5.1.1	Korrelationen zwischen Feinstaubpartikeln	10
5.1.2	Korrelationen zwischen Temperatur, Feuchtigkeit und Taupunkt	10
5.1.3	Korrelationen mit der Zeit	11
5.1.4	Nachverfolgung von Anwesenheit im Labor durch die Messdaten.....	12
5.1.5	Stromausfall am 17.04.2023 im Labor.....	13
5.1.6	Großbrand in Herzogenrath am 21.05.2023.....	14
5.2	Analyse der Messdaten eines Einpersonenhaushalts	15
5.2.1	Korrelationen zwischen Feinstaubpartikeln	17
5.2.2	Korrelation zwischen Temperatur und Schwefeloxid	17
5.2.3	Korrelation zwischen Temperatur und Sauerstoff	17
5.2.4	Korrelation zwischen Sauerstoff und Schwefeldioxid.....	18
5.2.5	Kochaktivitäten nicht klar abgrenzbar durch die Messdaten	19
5.2.6	Lüftungsvorgänge lassen sich klar in den Daten erkennen.....	20
5.2.7	Messfehler bzgl. Feinstaubwerte ausgelöst durch Luftfeuchtigkeit	21
5.3	Analyse des dritten Datensatzes.....	22
6.	Weiterführung und Ausbau des Projekts	23
6.1	Erweiterte Messungen	23
6.2	Implementierung eines Analyse-Dashboards.....	23
6.3	Automatisierte Datenanalyse und Erweiterung der Software-Funktionen	23
6.4	Anwendung von Machine Learning.....	24
6.5	Reengineering des airQ Sensors.....	24
7.	Fazit	25

1. Hintergrund und Ziel dieses Projekts

Das Ziel dieses Projekts war es, Daten über die Luftqualität zu sammeln und diese auszuwerten.

Die Betrachtungen konzentrierten sich dabei auf drei wesentliche Aspekte: Die Identifizierung von potenziellen Messfehlern, die Untersuchung von Korrelationen zwischen den verschiedenen erfassten Metriken und eine detailliertere Beurteilung der Luftqualität auf Basis von Korrelationen zu anderen Messwerten mit Betrachtung des auslösenden Ereignisses bei z.B. Grenzwertüberschreitungen einzelner Metriken.

2. Verwendete Sensorik

Zur Erhebung der Daten wurde der Luftanalyse-Sensor "air-Q science" von der Corant GmbH verwendet. Dieser bietet in der Standardversion 14 Sensoren zur Betrachtung von:

- Sauerstoff
- Kohlendioxid
- Kohlenmonoxid
- Feinstaub PM1 (Feinstaub mit einem Durchmesser von weniger als 1 Mikrometer)
- Feinstaub PM2,5 (Feinstaub mit einem Durchmesser von weniger als 2,5 Mikrometer)
- Feinstaub PM10 (Feinstaub mit einem Durchmesser von weniger als 10 Mikrometer)
- Stickstoffdioxid
- Schwefelwasserstoff
- Ozon
- VOC-Gase
- Temperatur
- Relative Luftfeuchte
- Luftdruck
- Lärm

Durch die oben genannte Metriken lassen sich noch diese Kennzahlen präzise berechnen:

- Absolute Luftfeuchte
- Taupunkt

Der air-Q Sensor führt alle Zwei Sekunden eine Messung durch. Diese Messungen werden durch den gleitenden Mittelwert aggregiert bis das Ergebnis nach 60 Messungen bzw. zwei Minuten in einer JSON-Datei auf der SD-Karte des air-Q Sensors persistent gespeichert wird. Die Ablage der JSON-Dateien befinden sich hierbei in der vorgegebenen Ordnerstruktur Jahr/Monat/Tag auf der internen SD-Karte.

Das Messintervall und die Aggregation lässt sich individuell anpassen, wurde jedoch für dieses Projekt bei den zuvor erwähnten Standardeinstellungen belassen.

Um auf die Daten zugreifen zu können bietet das Science-Modell eine API, die mittels HTTP-Requests angesprochen werden kann. Dazu muss eine WLAN-Verbindung von einem Endgerät aufgebaut werden oder der Sensor muss in das Heimnetzwerk integriert werden. Zudem bietet der Hersteller eigene Plattformen in Form von einer App und einer Website an, um die Daten betrachten zu können.

3. Technische Umsetzung

Um eine leichte Verfügbarkeit der historischen Daten zu schaffen, wurde eine Schnittstelle entwickelt, die die Daten des air-Q Sensors anfragt und in eine lokale Datenbank überführt. Somit sind die Daten für die computergestützte Analyse leicht zugänglich, entsprechend aufbereitet und immer verfügbar.

Um das Verwalten der Datenbank zu erleichtern, wurde ein Modul programmiert, dass die gängigsten Datenbank-Befehle abdeckt. Somit kann ein guter Workflow bei der Analyse gewährleistet werden, da man nicht zwischen den Programmiersprachen SQL und Python hin und her wechseln muss. Dies erleichtert auch das Verwalten der Datenbank für Nutzer ohne Vorkenntnisse.

Um auch die Analyse zu erleichtern, wurde ein eigenes Analyse-Modul programmiert. Dies stellt den Funktionsumfang bereit, um unerfahrenen Nutzern eine eigenständige Korrelationsanalyse zu ermöglichen und ohne große Hürde wichtige Visualisierungen zu erstellen.

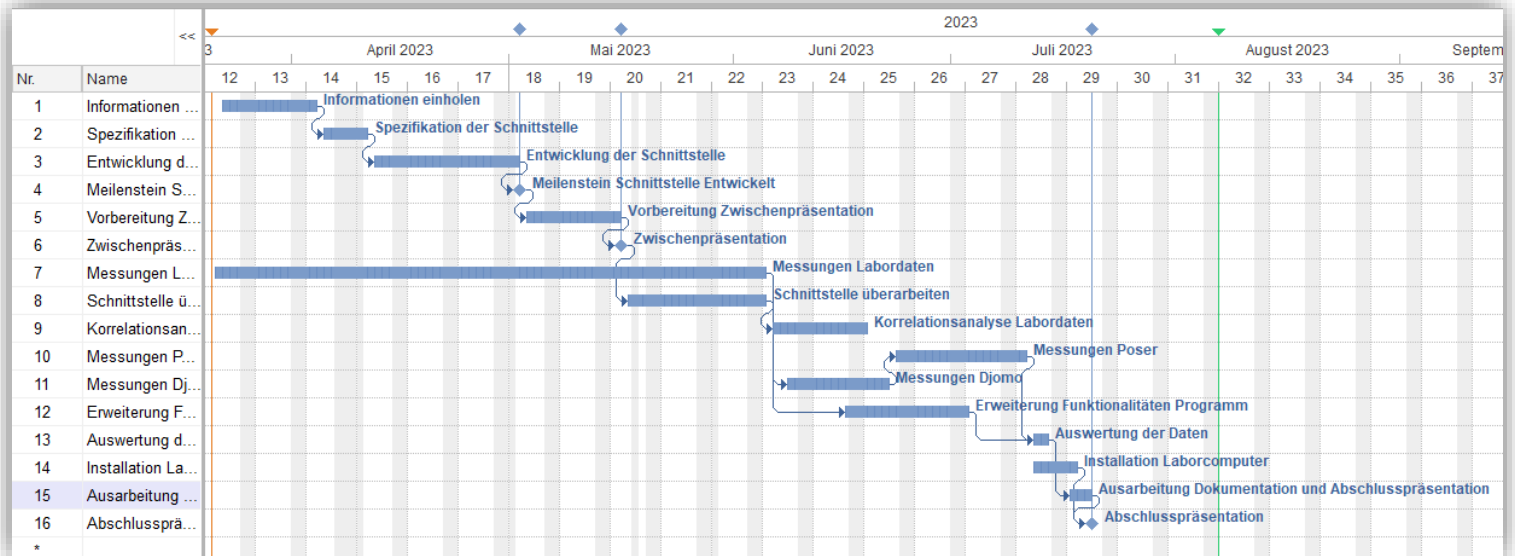
Der volle Funktionsumfang der beiden Module wurde mittels des Google-Style Docstring Formats dokumentiert. Diese Dokumentationsmethode kann bei Bedarf in eine API-Dokumentation umgewandelt werden. Dabei bieten sich Tools wie Sphinx oder Doxygen an. Durch diese Dokumentationsart ist eine effektive Weiterentwicklung und Fortführung des Projekts, auch ggf. durch andere Entwicklerteams, sichergestellt.

Um die Dokumentation zu ergänzen, wird noch ein Jupyter-Notebook namens "Anleitung.ipynb" bereitgestellt, indem alle Funktionen exemplarisch verwendet und beschrieben werden.

Tiefer wird in diesem Dokument nicht auf die technische Umsetzung eingegangen da diese Dokumentation einen allgemeinen Überblick und die Ergebnisse des Projekts erklären soll. Für mehr Details zur technischen Umsetzung empfiehlt es sich deshalb die Dokumentation der Module zu lesen und ggf. auch die bereitgestellte Anleitung zu konsultieren.

4. Projektablauf

In diesem Kapitel werden die Schritte des Datenanalyse-Projekts kurz erläutert. Dabei werden auf die durchgeführten Tätigkeiten eingegangen und ggf. auch auf aufgetretene Herausforderungen.



4.1 Informationen einholen

Um ein Verständnis über die funktionalen Möglichkeiten des air-Q Sensors zu erhalten, wurde ausgiebig die Dokumentation des Sensors studiert. Hierbei wurde sich vor allem über die Möglichkeiten informiert, wie effizient auf die Daten zugegriffen werden kann, ohne den Sensor in das Heimnetzwerk einbinden zu müssen. Zudem wurde sich über die Verschlüsselung und das Datenformat informiert.

4.2 Spezifikation der Schnittstelle

Auf Basis, der davor in Erfahrung gebrachten Informationen wurde, eine Lösung für die effiziente Datenhaltung erarbeitet und eine Schnittstelle konzipiert.

MariaDB wurde als Datenbank gewählt und installiert. Zudem wurde ein Schema für die Sensordaten entwickelt.

Für die Schnittstelle wurden folgende Kriterien identifiziert:

- Die Schnittstelle soll selbstständig die Datenbankverbindung aufbauen können
- Die verfügbaren Daten werden mittels einer Anfrage der bestehenden Ordnerstruktur ermittelt
- Die Anfrage muss alle verfügbaren Daten Anfragen können und diese selbstständig in der davor bereitgestellten Datenbank speichern
- Es dürfen keine Duplikate in der Datenbank vorhanden sein
- Die Datenbankverbindung wird vom Programm selbstständig, nach verrichteter Arbeit, wieder geschlossen.

4.3 Entwicklung der Schnittstelle

Die Entwicklung der Schnittstelle war nicht trivial. Das Iterieren der Ordnerstruktur ist komplex. Zudem gab es Konvertierungsprobleme mit dem vorgefertigten UNIX-Timestamp der in die Datenbank eingefügt wird.

Nach Entwicklung der Schnittstelle ist aufgefallen, dass die Anfrage der Sensordaten erheblich mehr Zeit in Anspruch nimmt als davor angenommen. Die Zeitverzögerung wird durch die http-Anfrage und die interne Verschlüsselung der http-Response des Sensors verursacht.

4.4 Schnittstelle Überarbeiten

Aufgrund der starken Zeitverzögerung bei der Datenanfrage musste die Schnittstelle überarbeitet werden. Während des Entwicklungsprozesses ist aufgefallen, dass die Schnittstelle zu diesem Zeitpunkt nur in der Lage war die Daten der letzten Zwei Monate anzufragen. Dies hatte mit einem internen Puffer, der nur Zwei Monate zwischenspeichert, zu tun der durch die Schnittstelle angefragt wurde.

Um beide Probleme zu lösen, wurde eine komplett neue Schnittstelle entwickelt. Diese fragt nicht mehr den Puffer an, sondern greift direkt auf die Daten, des zuvor definierten Zeitraums, zu. Zudem wird zuerst überprüft, ob sich schon Daten in der Datenbank befinden, die im anzufragenden Zeitraum liegen. Ist dies der Fall, dann wird die Datenanfrage erst ab den noch nicht in der Datenbank liegenden Daten ausgeführt. Dies kann erheblich Zeit sparen.

4.5 Korrelationsanalyse Labordaten

Die Korrelationsanalyse verlief, nach kurzer Einarbeitung in die Funktionalitäten von Pandas, problemlos. Jedoch wurde bei der Durchführung der Korrelationsanalyse schon der Bedarf für weitere Funktionalitäten zwecks der Datenbank und der Analyse festgestellt.

Die Ergebnisse der Datenanalyse werden in den nachfolgenden Kapiteln erläutert.

4.6 Erweiterung Funktionalität Programm

Um eine unkomplizierte Datenanalyse mit zugehöriger Datenbankverwaltung zu erreichen, wurden verschiedene Funktionen entwickelt.

Nach der Entwicklung der Funktionen wurde entschieden die ganzen Funktionen objektorientiert zu kapseln, um die Übersichtlichkeit und Wiederverwendbarkeit zu steigern und sicherzustellen.

Diese Kapselung zusammen mit der flächendeckenden Dokumentation durch Docstrings war ein hoher zeitlicher Aufwand.

4.7 Auswertung der Daten

Die Analyse der Daten lief weitestgehend problemlos. Für die Analyse wurden die eigens kreierten Module und Microsoft Power BI verwendet. Die Ergebnisse der Datenanalyse werden im folgenden Kapitel erläutert.

4.8 Installation Laborcomputer

Für die Installation der Programme auf den Laborrechnern der FH Aachen wurde die Hilfe der DVZ benötigt. Ohne Administrationsrechte ist man nicht in der Lage verschiedene Programme zu installieren und den WLAN-Stick für den Standrechner einzurichten. Dieser wird zwingend gebraucht, um die WLAN-Verbindung zum airQ Sensor aufzubauen.

5. Analyseergebnisse der Daten

In diesem Kapitel werden die Analyseergebnisse der ausgewerteten Messdaten vorgestellt.

Um ausreichend Varianz in den Daten zu haben, wurde an drei unterschiedlichen Orten Messungen durchgeführt. Die Messorte waren ein Elektrotechniklabor in der FH Aachen, an dem an Radaranlagen geforscht wird und zwei kleine private Haushalte.

Die Analyse des Korrelationsverhaltens zwischen den verschiedenen Metriken erfolgt in mehreren sorgfältig geplanten Schritten. Zunächst wird eine mathematische Korrelationsanalyse auf grober Ebene durchgeführt. Diese wird mit der Hilfe der `'corr'` Methode von pandas umgesetzt, die den Korrelationskoeffizienten zwischen allen Paaren von numerischen Variablen in dem DataFrame berechnet. Der Standard dabei ist der Pearson-Korrelationskoeffizient, der die lineare Abhängigkeit zwischen zwei Variablen misst und Werte zwischen -1 und 1 liefert. Dies gibt einen ersten Überblick über mögliche Korrelationen in den Daten.

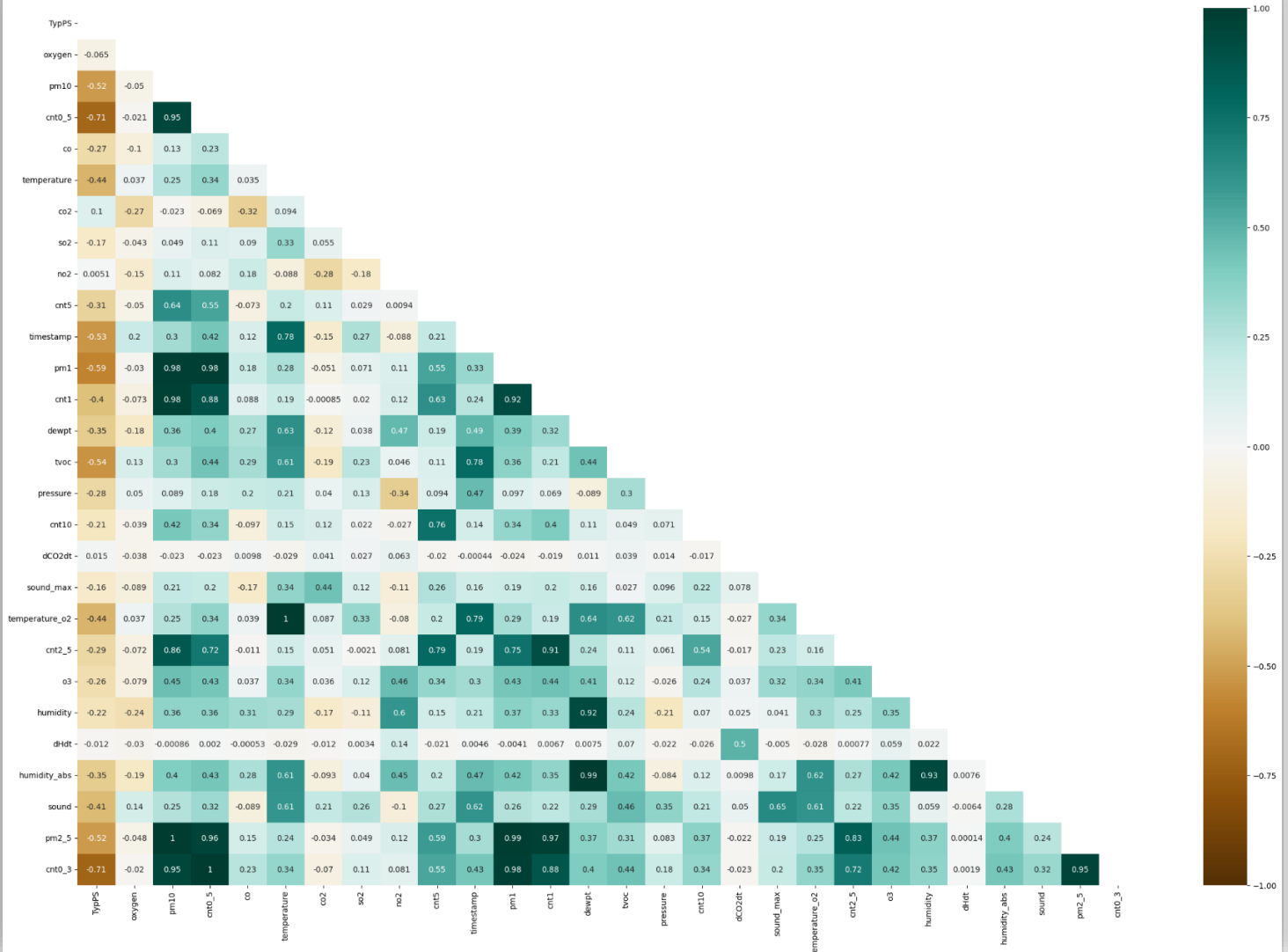
Nach diesem ersten Schritt erfolgt eine detailliertere Untersuchung der Korrelationen durch die Erstellung von Streudiagrammen oder Scatter Plots. Dies ist ein wichtiges Element in der Datenanalyse, da es eine visuelle Bestätigung der statistisch ermittelten Korrelationen ermöglicht. Scatter Plots bieten die Möglichkeit, Muster, Trends und Ausreißer in den Daten aufzudecken und diese Korrelationen visuell zu validieren. Besonders wertvoll ist hier die Möglichkeit, optional eine Trendlinie in den Scatter Plot zu integrieren, um so den allgemeinen Trend der Datenpunkte besser sichtbar zu machen.

Im letzten Schritt der Analyse wird der zeitliche Verlauf der korrelierten Metriken im Detail anhand von Liniendiagrammen untersucht. Hierbei wird nicht nur die reine Korrelation berücksichtigt, sondern auch versucht, diese im Kontext der Messung und der gegebenen Umstände zu interpretieren. Hier fließen wissenschaftliche Erkenntnisse ein, aber auch eine genaue Kenntnis der Bedingungen während der Messung ist hier sehr hilfreich. So wird versucht, ein vollständiges Bild der Situation zu zeichnen und nicht nur reine Zahlen zu betrachten.

Die Kombination dieser Schritte ermöglicht eine umfassende Analyse und Interpretation der Daten. Es ist ein Ansatz, der sowohl auf harten Zahlen als auch auf visueller Darstellung und Einbeziehung von Kontextinformationen basiert.

5.1 Analyse der Messdaten des Labors

Triangle Correlation Heatmap

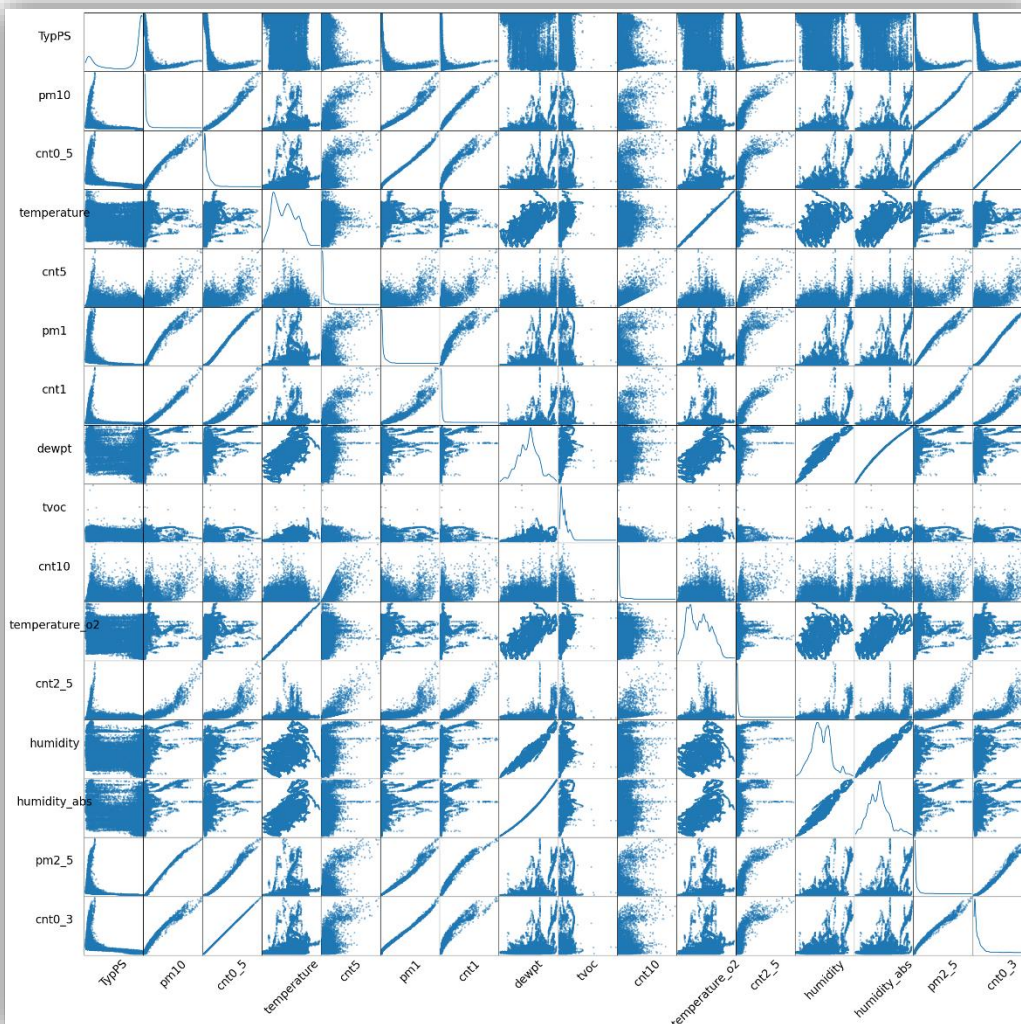


Auffällig beim Datensatz des Labors ist, dass nur zu erwartende Korrelationen entdeckt worden sind. Alle ermittelten Korrelationen lassen sich plausibel erklären. Dies ist jedoch zu erwarten, da über einen langen Zeitraum gemessen wurde und somit kleine Anomalien in der Masse der Daten untergehen. Zudem finden in diesem Labor wenig Aktivitäten statt die die Metriken in einem Erkennbaren Maß beeinflussen.

Folgende Korrelationen wurden bei einem Grenzwert des Korrelationsindex von größer 0.7, durch das zuvor erstellte Programm, ermittelt:

Correlation between cnt0_5 and cnt0_3:	0.9998919837574829
Correlation between temperature_o2 and temperature:	0.9991241583966602
Correlation between pm10 and pm2_5:	0.9965343875194026
Correlation between humidity_abs and dewpt:	0.993294258144139
Correlation between pm1 and pm2_5:	0.9862348986898426
Correlation between cnt0_5 and pm1:	0.9785995476375734
Correlation between pm1 and pm10:	0.9779836573855784

Correlation between pm1 and cnt0_3:	0.9772604176877829
Correlation between pm10 and cnt1:	0.9761367377385244
Correlation between cnt1 and pm2_5:	0.9693201992950952
Correlation between cnt0_5 and pm2_5:	0.9568793051180247
Correlation between pm2_5 and cnt0_3:	0.9546523627609094
Correlation between cnt0_5 and pm10:	0.9517993119057035
Correlation between pm10 and cnt0_3:	0.9494665059775276
Correlation between humidity_abs and humidity:	0.9310335587114708
Correlation between dewpt and humidity:	0.9238826361175861
Correlation between pm1 and cnt1:	0.9192948007346963
Correlation between cnt1 and cnt2_5:	0.9077935727407743
Correlation between cnt0_5 and cnt1:	0.8833790866058397
Correlation between cnt1 and cnt0_3:	0.8799711011213327
Correlation between pm10 and cnt2_5:	0.8623609248762336
Correlation between pm2_5 and cnt2_5:	0.8294332559550949
Correlation between timestamp and temperature_o2:	0.7928244319389156
Correlation between cnt5 and cnt2_5:	0.7856012729677331
Correlation between timestamp and temperature:	0.7814409326884955
Correlation between timestamp and tvoc:	0.7807787661154336
Correlation between cnt10 and cnt5:	0.7553587562098848
Correlation between pm1 and cnt2_5:	0.7498457753315652
Correlation between cnt0_5 and cnt2_5:	0.7220030964463633
Correlation between cnt2_5 and cnt0_3:	0.7181754392650567
Correlation between TypPS and cnt0_3:	-0.713258780180892
Correlation between cnt0_5 and TypPS:	-0.709347774802263



5.1.1 Korrelationen zwischen Feinstaubpartikeln

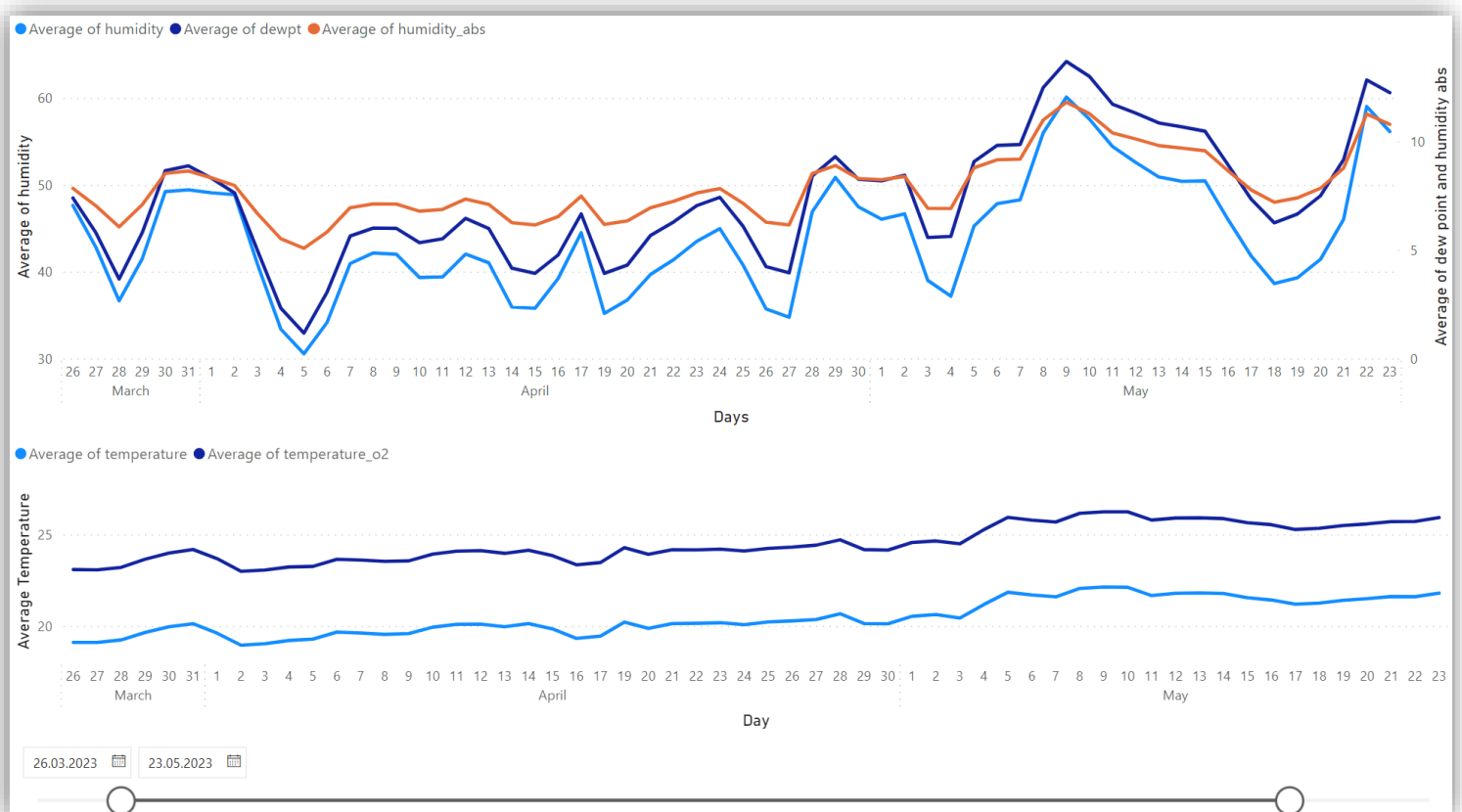
Unsere Korrelationsanalyse offenbart signifikante positive Korrelationen zwischen verschiedenen Metriken für Feinstaubpartikel, die uns Aufschluss über die Anzahl und Größe der in der Luft vorhandenen Partikel geben. Bei einer näheren Betrachtung dieser Korrelationen ist es wichtig, die Unterschiede und den Zusammenhang zwischen den zwei Schlüsselmetriken – CNT (Count Metrics) und PM (Particulate Matter) – zu verstehen.

CNT repräsentiert die Anzahl von Partikeln in der Luft, die größer als eine bestimmte Größe sind. Bei einer hohen Korrelation zwischen den Messwerten von CNT0.5 und CNT0.3 – beispielsweise ca. 0,9999 in unseren Daten – bedeutet dies, wenn die Anzahl der Partikel, die größer als 0,5 Mikrometer sind, ansteigt, auch die Anzahl der Partikel, die größer als 0,3 Mikrometer sind, tendenziell ansteigt. Dies ist logisch, da die Partikel mit einem Durchmesser von 0,5 Mikrometer in der Gesamtzahl der Partikel mit einem Durchmesser von 0,3 Mikrometer enthalten sind.

PM hingegen ist eine Metrik für die Masse der Partikel in der Luft, die kleiner als eine bestimmte Größe sind. Wenn wir eine hohe Korrelation zwischen PM1 und PM2.5 von etwa 0,986 feststellen, dann bedeutet das, wenn die Masse der Partikel, die kleiner als 1 Mikrometer sind, zunimmt, auch die Masse der Partikel, die kleiner als 2,5 Mikrometer sind, tendenziell ansteigt. Ähnlich wie bei den CNT-Werten ist dies verständlich, da die kleineren Partikel (PM1) in der Masse der größeren Partikel (PM2.5) enthalten sind.

5.1.2 Korrelationen zwischen Temperatur, Feuchtigkeit und Taupunkt

In unserer Korrelationsanalyse haben wir eine starke positive Korrelation zwischen den Messwerten für Temperatur, Feuchtigkeit und Taupunkt festgestellt. Diese eng verwobenen Beziehungen ergeben sich aus der physikalischen Dynamik dieser Faktoren in der Atmosphäre.



Die Temperatur hat einen direkten Einfluss auf die Menge an Wasser, die die Luft aufnehmen kann, wobei wärmere Luft in der Lage ist, mehr Feuchtigkeit zu halten. Wenn die Lufttemperatur steigt, nimmt auch ihre Kapazität zur Aufnahme von Feuchtigkeit zu, was zu einem Anstieg der absoluten Luftfeuchtigkeit führt.

Die relative Luftfeuchtigkeit, die in Prozent gemessen wird, gibt an, wie viel Feuchtigkeit die Luft im Verhältnis zur maximal möglichen Menge bei der aktuellen Temperatur enthält. Daher kann sich die relative Luftfeuchtigkeit auch ohne Veränderung der absoluten Luftfeuchtigkeit ändern, wenn die Temperatur variiert.

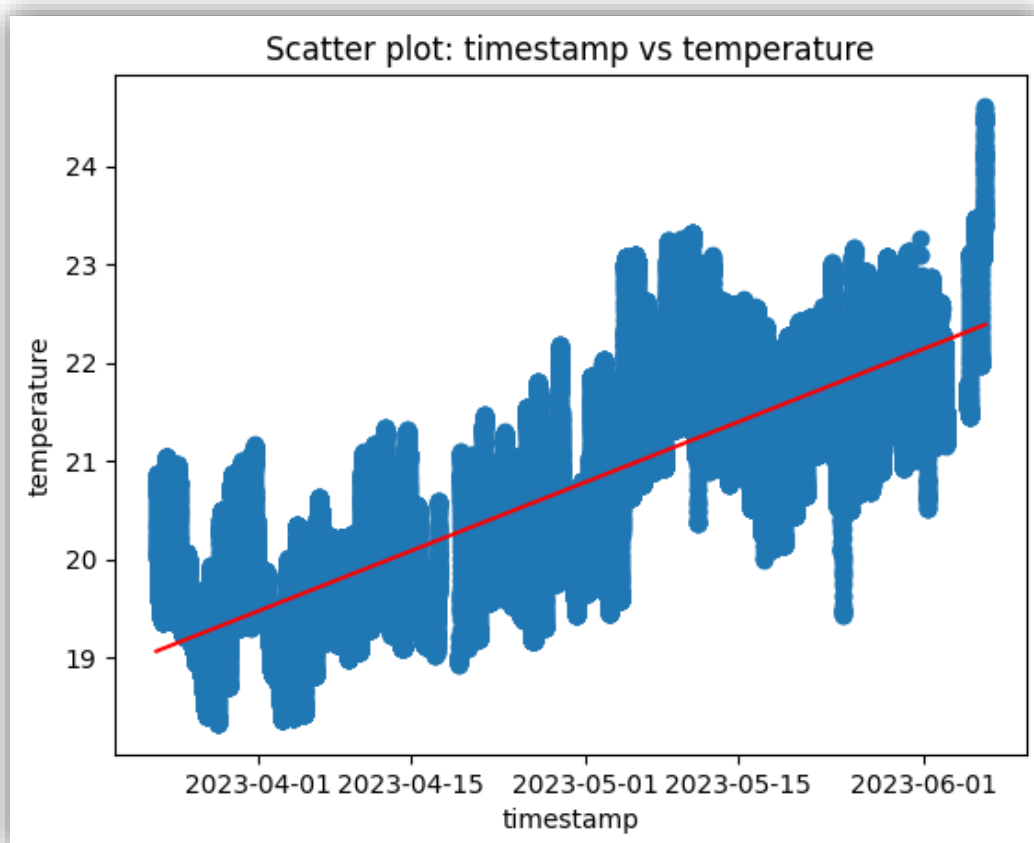
Der Taupunkt ist die Temperatur, bei der die Luft gesättigt ist und Wasser zu kondensieren beginnt. Dies geschieht, wenn die Luft abkühlt und ihre Fähigkeit zur Aufnahme von Feuchtigkeit abnimmt. Daher ist der Taupunkt direkt mit der absoluten Feuchtigkeit in der Luft verbunden und steigt mit zunehmender Feuchtigkeit. Diese Beziehung erklärt die starke positive Korrelation (0,993) zwischen der absoluten Feuchtigkeit und dem Taupunkt in unseren Daten.

5.1.3 Korrelationen mit der Zeit

Die Analyse hat ebenfalls gezeigt, dass einige Metriken eine signifikante Korrelation mit der Zeit aufweisen. Insbesondere die Temperaturmessungen zeigen eine positive Korrelation von etwa 0,793 (für `temperature_o2`) und 0,781 (für `temperature`) mit der Zeit.

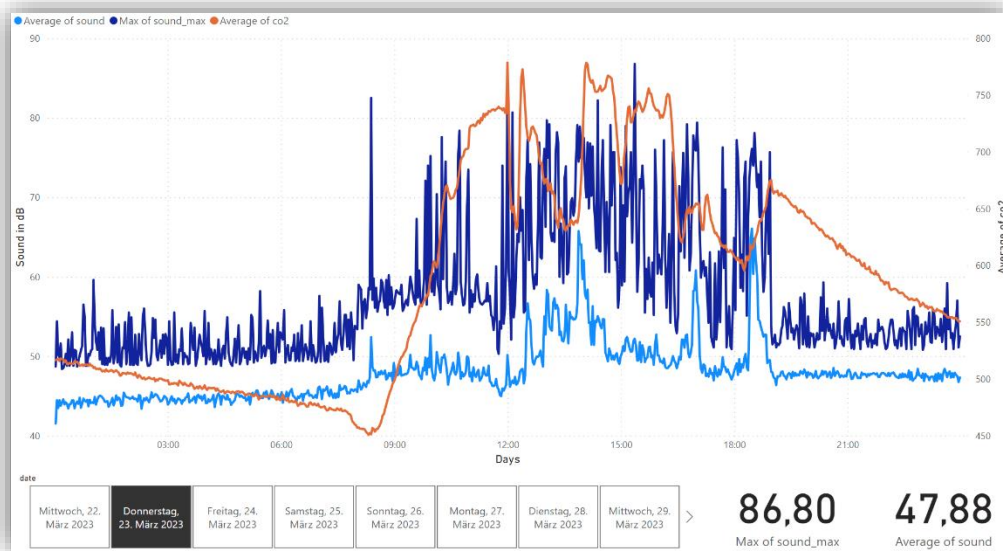
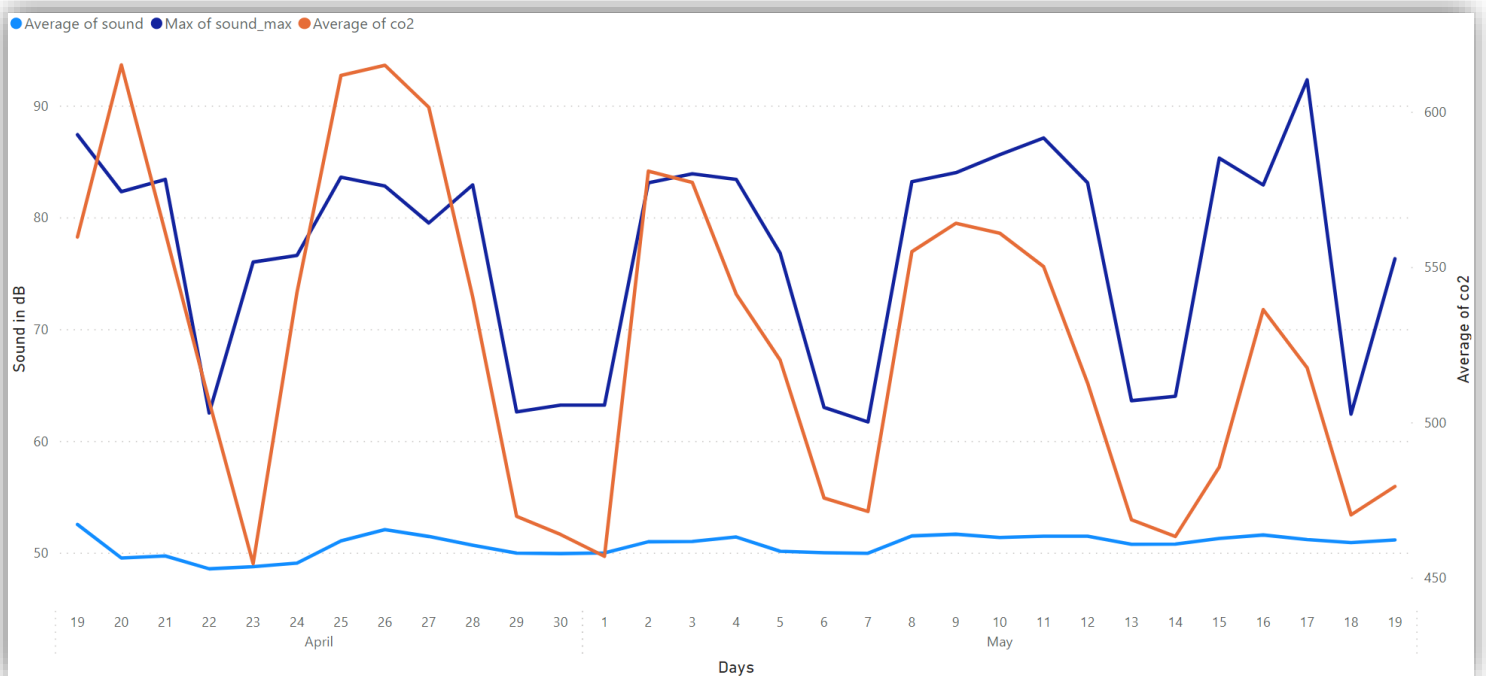
Es ist wichtig zu beachten, dass die Messungen über eine Zeitspanne vom Frühling bis zum Sommerbeginn erfasst wurden. Während dieser Zeit steigen die Temperaturen im Allgemeinen aufgrund der jahreszeitlichen Veränderungen. Dies erklärt die positive Korrelation zwischen Temperatur und Zeit.

Zusätzlich können tägliche Temperaturschwankungen zu dieser Korrelation beitragen. Typischerweise sinken die Temperaturen in der Nacht und steigen tagsüber, was eine zyklische Abhängigkeit der Temperatur von der Tageszeit erzeugt.



5.1.4 Nachverfolgung von Anwesenheit im Labor durch die Messdaten

Am Verlauf der Metriken des maximal gemessenen Lärmpegels und der durchschnittlichen co2-Konzentration lässt sich ableiten ob und wann sich Personen im Labor befinden. Dabei ist ein klar zyklischer Verlauf der beiden Metriken zu erkennen. Beide Metriken fallen an Wochenenden ab da dort das Labor in der Regel nicht zugänglich ist.

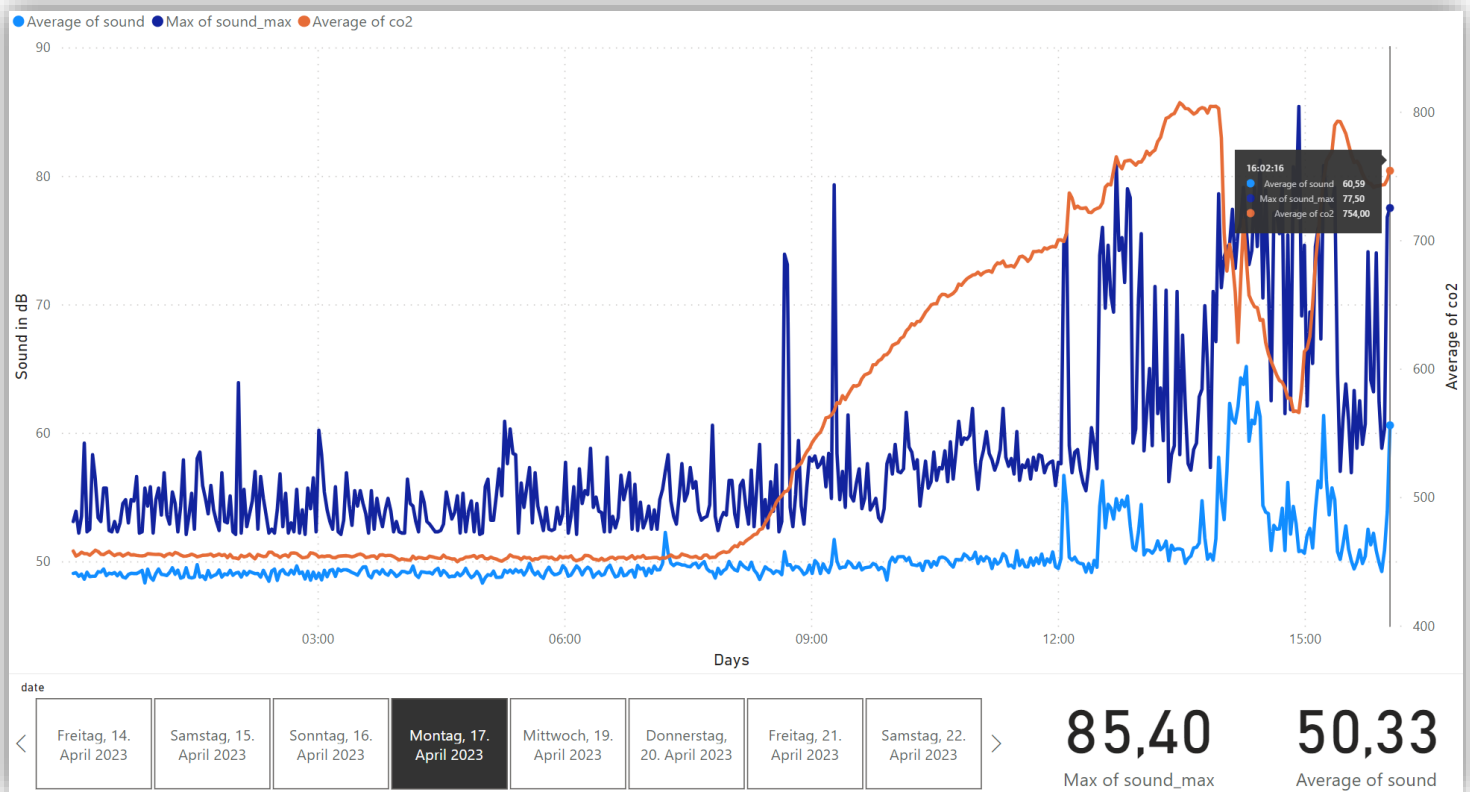


Zudem lässt sich, bei feingranularer Betrachtung eines einzelnen Tages, relativ genau die Uhrzeiten feststellen, in denen das Labor genutzt wird.

Der sporadische Abfall der co2-Konzentration lässt auf ein geöffnetes Fenster oder eine geöffnete Tür hindeuten.

5.1.5 Stromausfall am 17.04.2023 im Labor

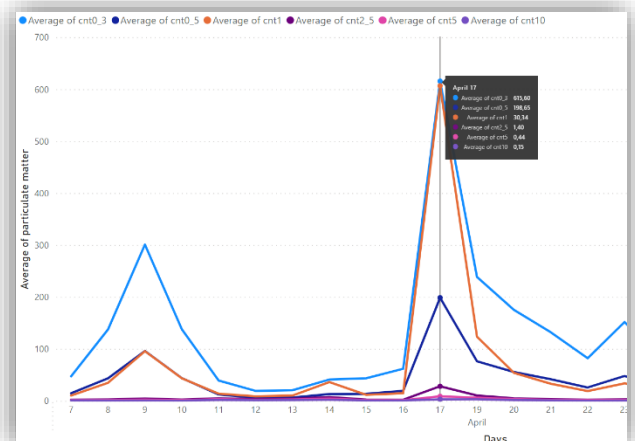
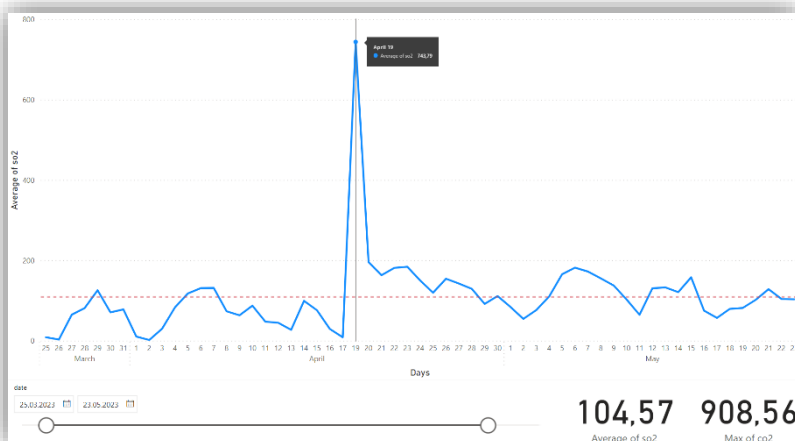
Am Montag, den 17.04.2023 gab es ab 16:02 einen Stromausfall im Labor. Die genaue Uhrzeit ist festzustellen durch den Abbruch der kontinuierlichen Messwerte.



Interessant dabei ist, dass die Messwerte darauf hindeuten, dass sich während des Stromausfalls Personen im Labor befunden haben müssen. Jedoch wurde kein Vorfall den Verantwortlichen des Labors gemeldet.

Darüber hinaus ist, nach wieder Inbetriebnahme des Sensors, die höchste Konzentration von Schwefeldioxid in der Luft festzustellen, die über den gesamten Messverlauf der Daten erhoben wurde.

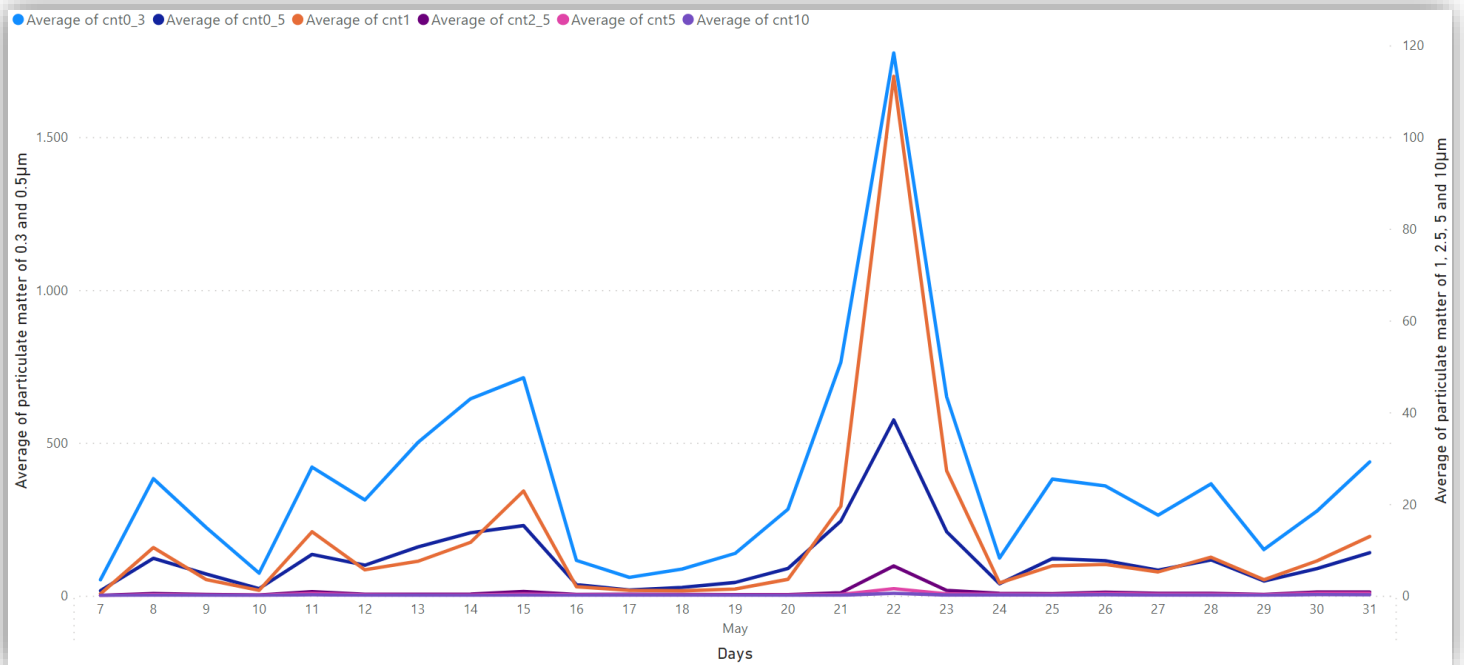
Die hohe Konzentration von Schwefeldioxid und ein gleichzeitiger Anstieg von Feinstaubpartikeln kann auf einen Verbrennungsvorgang hindeuten. Möglicherweise ist in dem Elektrotechnik-Labor ein Kurzschluss entstanden der einen Stromausfall, durch eine ausgelöste Sicherung, verursacht hat und durch die entstandene Verbrennung die Luftwerte dementsprechend beeinflusst hat.



5.1.6 Großbrand in Herzogenrath am 21.05.2023

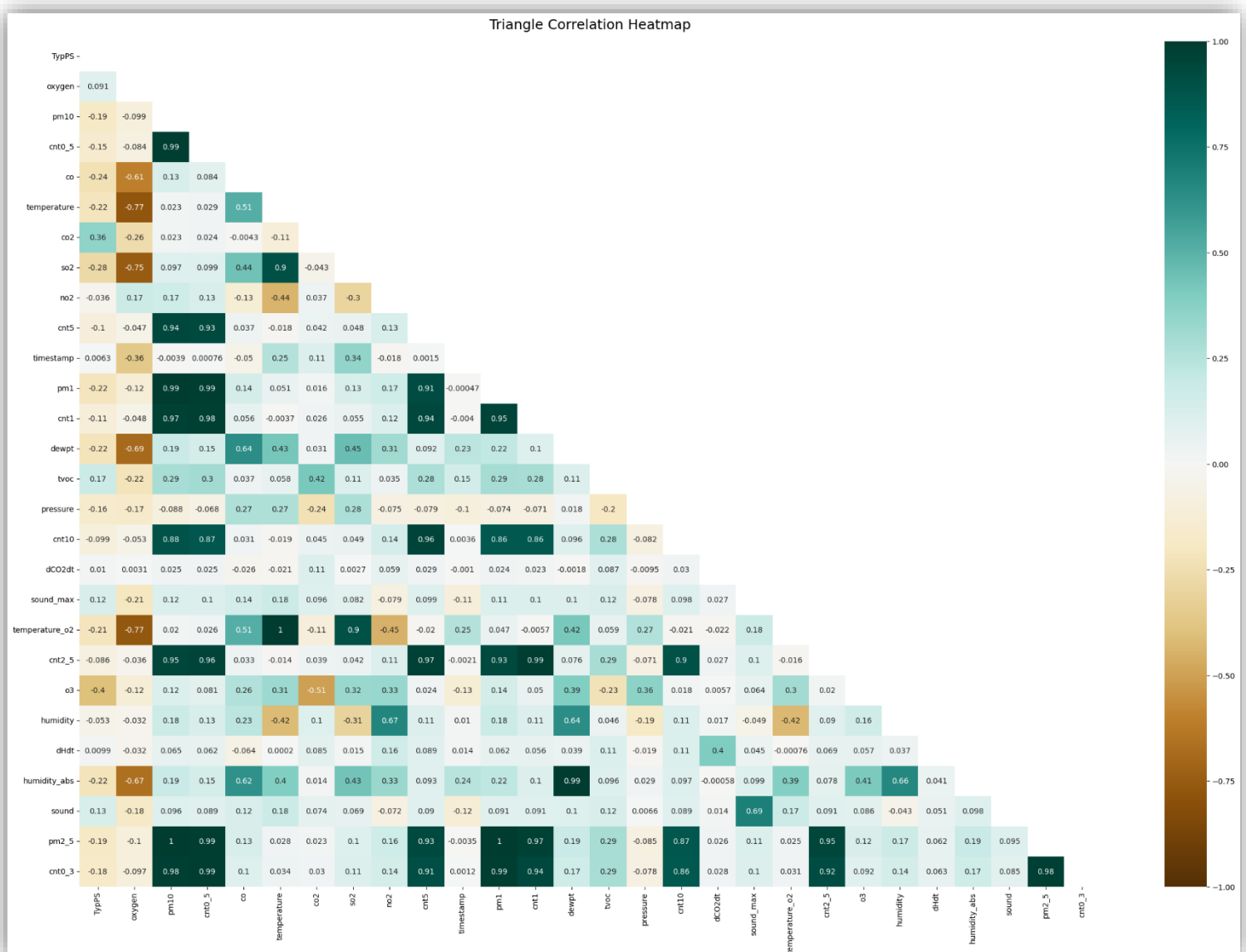
Am Abend des 21.05.2023 wurde die Feuerwehr wegen starker Rauchentwicklung, die vom ehemaligen Verwaltungsgebäude des Eschweiler Bergwerksvereins ausging, alarmiert. ([Quelle](#))

Obwohl die Messstelle ca. 9-10 Kilometer vom Feuer entfernt war, kann man die Auswirkungen des Feuers für die Luftqualität deutlich in den Messergebnissen erkennen.



Es ist ein deutlicher Anstieg der Feinstaubwerte während des Zeitraums des Feuers zu sehen. Gerade deshalb wurde auch über die Warnapp NINA gemeldet, dass man im Kreis Aachen seine Fenster über einen gewissen Zeitraum geschlossen halten soll.

5.2 Analyse der Messdaten eines Einpersonenhaushalts

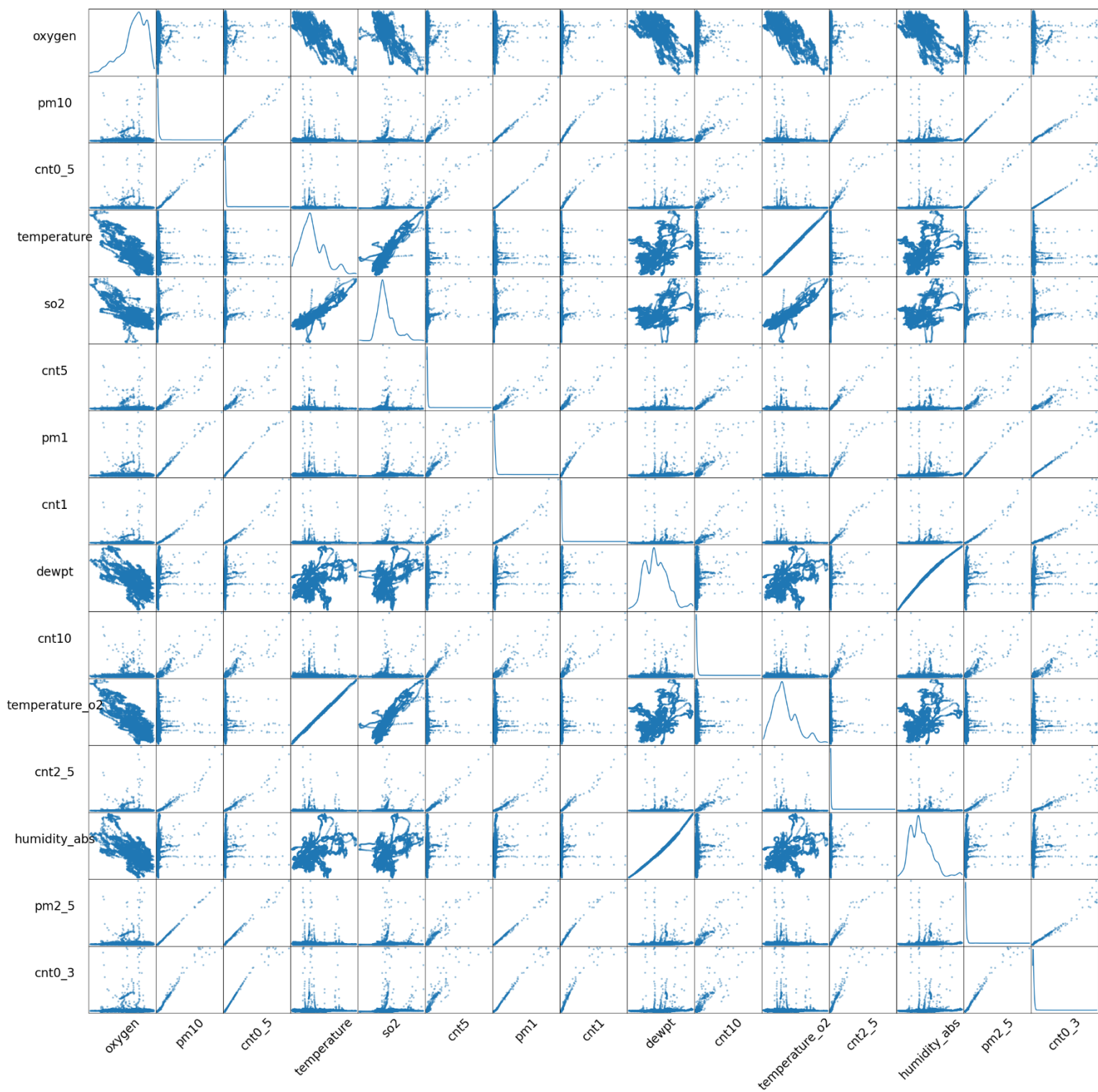


Um keine Korrelationen doppelt zu erklären werden in diesem Kapitel nur die Korrelationen beschrieben die einzigartig in diesem Erhebungszeitraum, im Vergleich zu den Labordaten, entdeckt wurden.

Einzigartige Korrelationen Ein-Zimmer-Wohnung (Ein-Personen-Haushalt):

Correlation between cnt1 and cnt10: 0.8607145334042409
 Correlation between cnt10 and cnt0_3: 0.8621656118555924
 Correlation between temperature_o2 and so2: 0.895740064461331
 Correlation between pm1 and cnt10: 0.8614293598382486
 Correlation between cnt1 and cnt5: 0.9448945661052653
 Correlation between pm2_5 and cnt5: 0.9327270775609914
 Correlation between cnt5 and cnt0_3: 0.9077322003835995
 Correlation between pm10 and cnt10: 0.8837561526406771
 Correlation between pm10 and cnt5: 0.939621211055441
 Correlation between so2 and oxygen: -0.7506743153617469
 Correlation between cnt0_5 and cnt10: 0.8710201458852372
 Correlation between cnt5 and pm1: 0.9140875095248342

Correlation between temperature and oxygen: -0.7739669011560055
 Correlation between pm2_5 and cnt10: 0.8726855857570549
 Correlation between so2 and temperature: 0.9002702705687103
 Correlation between cnt2_5 and cnt10: 0.897979644518529
 Correlation between cnt0_5 and cnt5: 0.9311642976540917
 Correlation between temperature_o2 and oxygen: -0.7719729173879115



5.2.1 Korrelationen zwischen Feinstaubpartikeln

Auffällig ist, dass im Vergleich zum Datensatz des Labors einige einzigartige Feinstaubkorrelationen festzustellen sind. Die Wohnung der Messung befindet sich ca. 80 Meter von einer Hauptstraße entfernt und ca. 120 Meter von dem Aachener Hauptbahnhof. Möglicherweise enthält die Luft deshalb eine andere Feinstaubzusammensetzung wie in den Labordaten.

Des Weiteren kann es sein, dass andere Feinstaubpartikel durch Vorgänge wie kochen produziert werden.

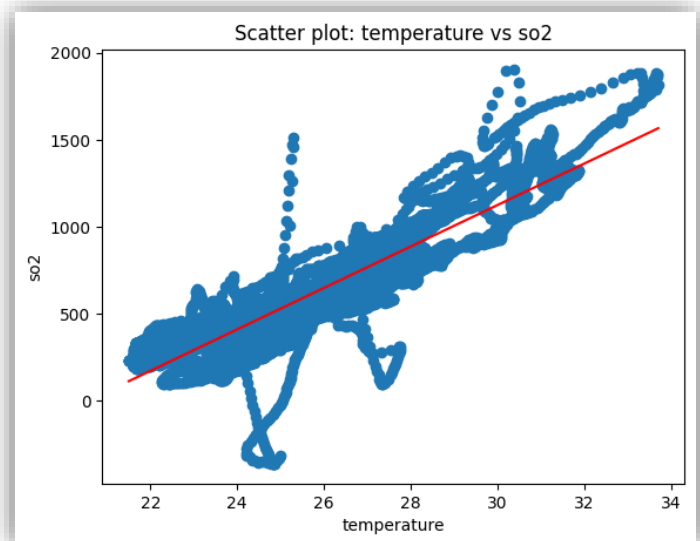
Zudem ist zu beachten, dass die Daten in einer Dachgeschosswohnung im Sommer erhoben worden sind. Möglicherweise kann durch eine stärkere thermische Aktivität, ausgelöst durch die Hitze in der Dachgeschosswohnung, mehr Feinstaubpartikel aufgewirbelt werden.

5.2.2 Korrelation zwischen Temperatur und Schwefeldioxid

Es gibt eine hohe positive Korrelation zwischen Temperatur und Schwefeldioxid (0.9). Diese Korrelation könnte durch externe Faktoren verursacht werden. Schwefeldioxid ist ein bekanntes Nebenprodukt der Verbrennung fossiler Brennstoffe. Es könnte sein, dass mit steigender Temperatur mehr Schwefeldioxid aus externen Quellen in die Wohnung eindringt. Dies könnte insbesondere dann der Fall sein, wenn die Temperaturerhöhung dazu führt, dass Fenster geöffnet werden, oder wenn es eine größere Nutzung von fossilbetriebenen Fahrzeugen oder Industrieanlagen in der Nähe gibt.

Ein weiterer Aspekt ist, dass die Diffusion von Gasen mit zunehmender Temperatur zunimmt, was zu einer erhöhten Verteilung von SO₂ in der Umgebung führen kann.

In Bezug auf die spezifische Situation in der Wohnung könnte diese hohe Korrelation auch durch das Kochen erklärt werden. Wenn schwefelhaltige Lebensmittel gekocht werden, wie zum Beispiel bestimmte Gemüsesorten, könnte dies zu einer erhöhten Freisetzung von SO₂ führen, besonders wenn dies bei hoher Temperatur geschieht.



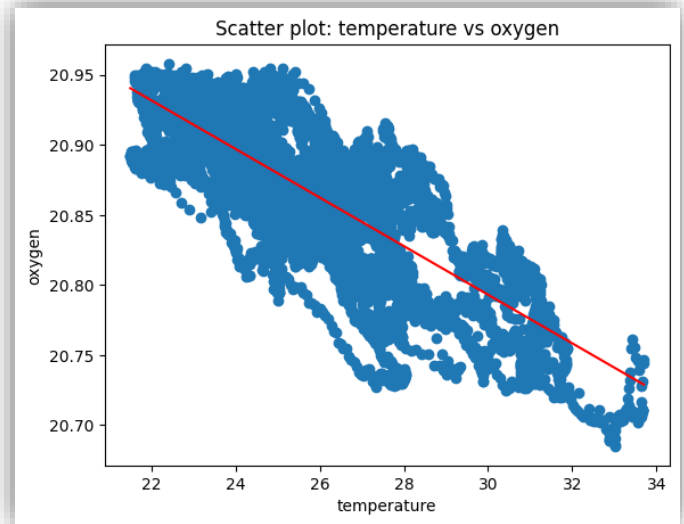
5.2.3 Korrelation zwischen Temperatur und Sauerstoff

Die beobachtete negative Korrelation zwischen Temperatur und Sauerstoff (-0.774) kann durch einige wissenschaftliche Grundsätze erklärt werden.

Einer der Gründe für diese Korrelation liegt in den grundlegenden physikalischen Eigenschaften von Gasen. Sauerstoff, wie alle Gase, folgt der allgemeinen Gasgesetzgleichung. Diese besagt, dass das Volumen eines Gases bei konstantem Druck mit steigender Temperatur zunimmt. Im Kontext der Atmosphäre bedeutet das, wenn die Temperatur steigt, sich das Volumen des Sauerstoffs ausdehnt. Da die Menge an Sauerstoff in einem bestimmten Raumvolumen konstant bleibt, führt die Volumenexpansion dazu, dass die Dichte des Sauerstoffs abnimmt. In der Praxis würde dies eine

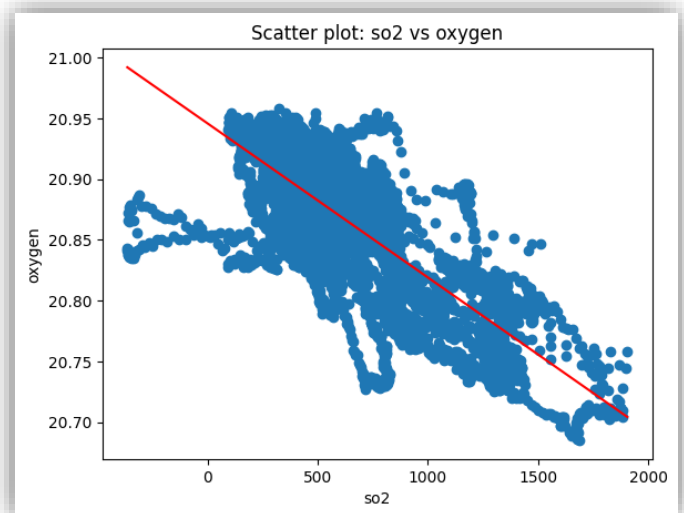
Abnahme der Sauerstoffkonzentration bedeuten, da die Menge an Sauerstoff pro Volumeneinheit abnimmt, wenn die Temperatur steigt.

Zudem ist es wichtig zu erwähnen, dass sich Temperaturänderungen auch auf biologische Prozesse auswirken können, die sich indirekt auf den Sauerstoffgehalt in der Luft auswirken. Bei wärmeren Temperaturen tendieren Organismen, einschließlich des Menschen, dazu, schneller und tiefer zu atmen, um ihre Körpertemperatur zu regulieren und einen kühlen Zustand aufrechtzuerhalten. Dies kann in Innenräumen zu einer höheren Sauerstoffaufnahme führen, wodurch der Sauerstoffgehalt in der Umgebungsluft sinkt.



5.2.4 Korrelation zwischen Sauerstoff und Schwefeldioxid

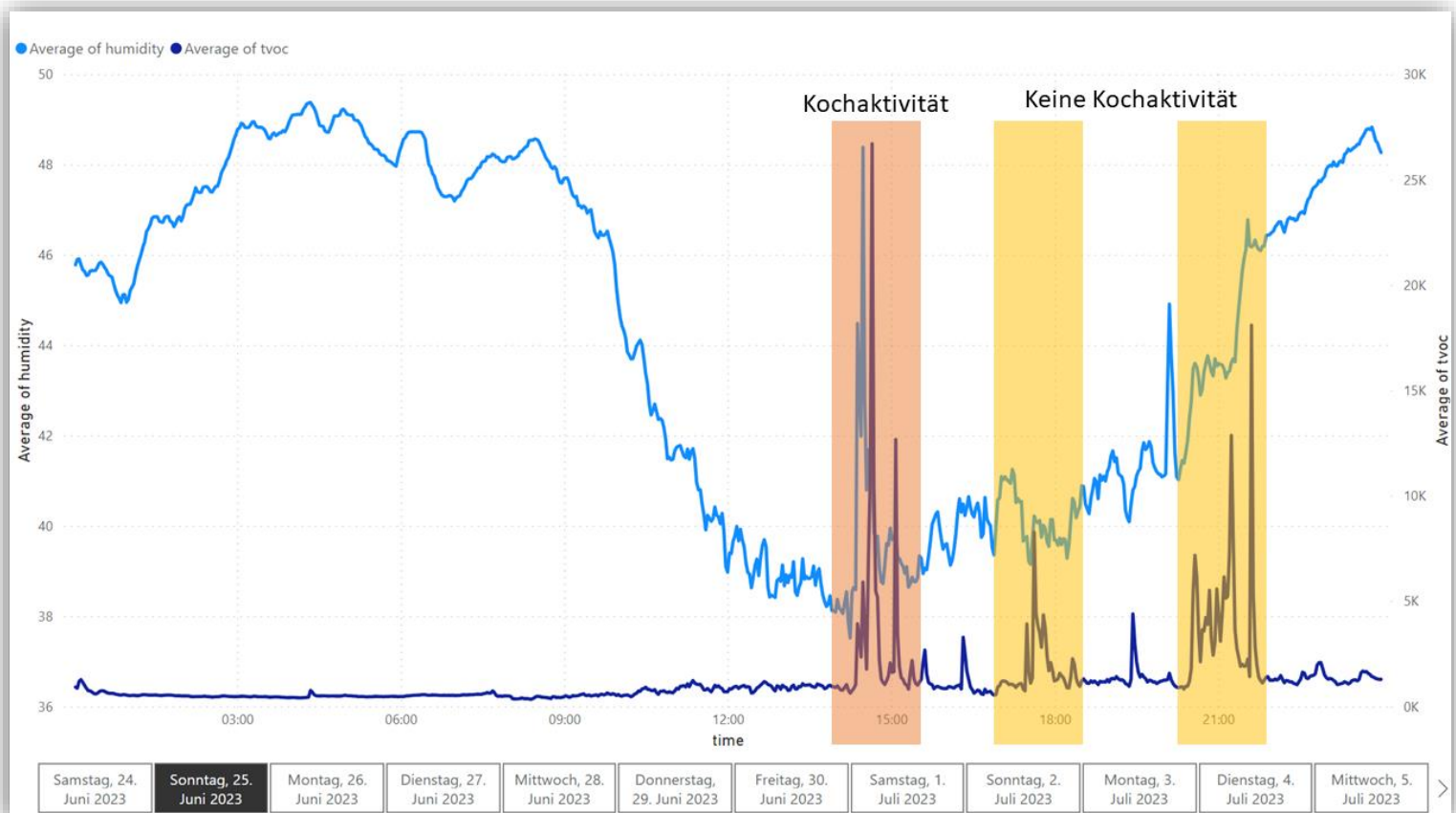
Die negative Korrelation zwischen Sauerstoff und Schwefeldioxid (-0.751) ist ein Indikator dafür, dass mit steigender SO₂-Konzentration die Sauerstoffkonzentration abnimmt. Eine mögliche wissenschaftliche Erklärung könnte die chemische Reaktion zwischen SO₂ und O₂ zu Schwefeltrioxid (SO₃) sein, die bei Anwesenheit eines Katalysators abläuft. In Bezug auf die häusliche Umgebung könnte dies durch Kochprozesse hervorgerufen werden, bei denen schwefelhaltige Lebensmittel gekocht werden, was zur Freisetzung von SO₂ führt und gleichzeitig Sauerstoff verbraucht.



5.2.5 Kochaktivitäten nicht klar abgrenzbar durch die Messdaten

Im Zuge unserer Analysen versuchten wir, Kochaktivitäten anhand des Verlaufs der Luftfeuchtigkeit (`humidity`) und der flüchtigen organischen Verbindungen (`VOC`) zu identifizieren. Allerdings stellten wir fest, dass sowohl die Luftfeuchtigkeit als auch die VOC-Werte hohe Volatilität aufweisen. Vor allem der VOC-Wert weist eine extreme Sensibilität auf wodurch oft Sprünge ohne nachvollziehbaren Grund in den Werten zu beobachten sind. Diese Eigenschaften erschwerten es uns, verlässliche Muster in den Kurvenverläufen zu identifizieren, die eindeutig auf Kochaktivitäten hindeuten würden. Jedoch ist festzustellen, dass ein Kochvorgang in der Regel einen zeitgleich sprunghaften Anstieg der Luftfeuchtigkeit und der VOC-Werte nach sich zieht.

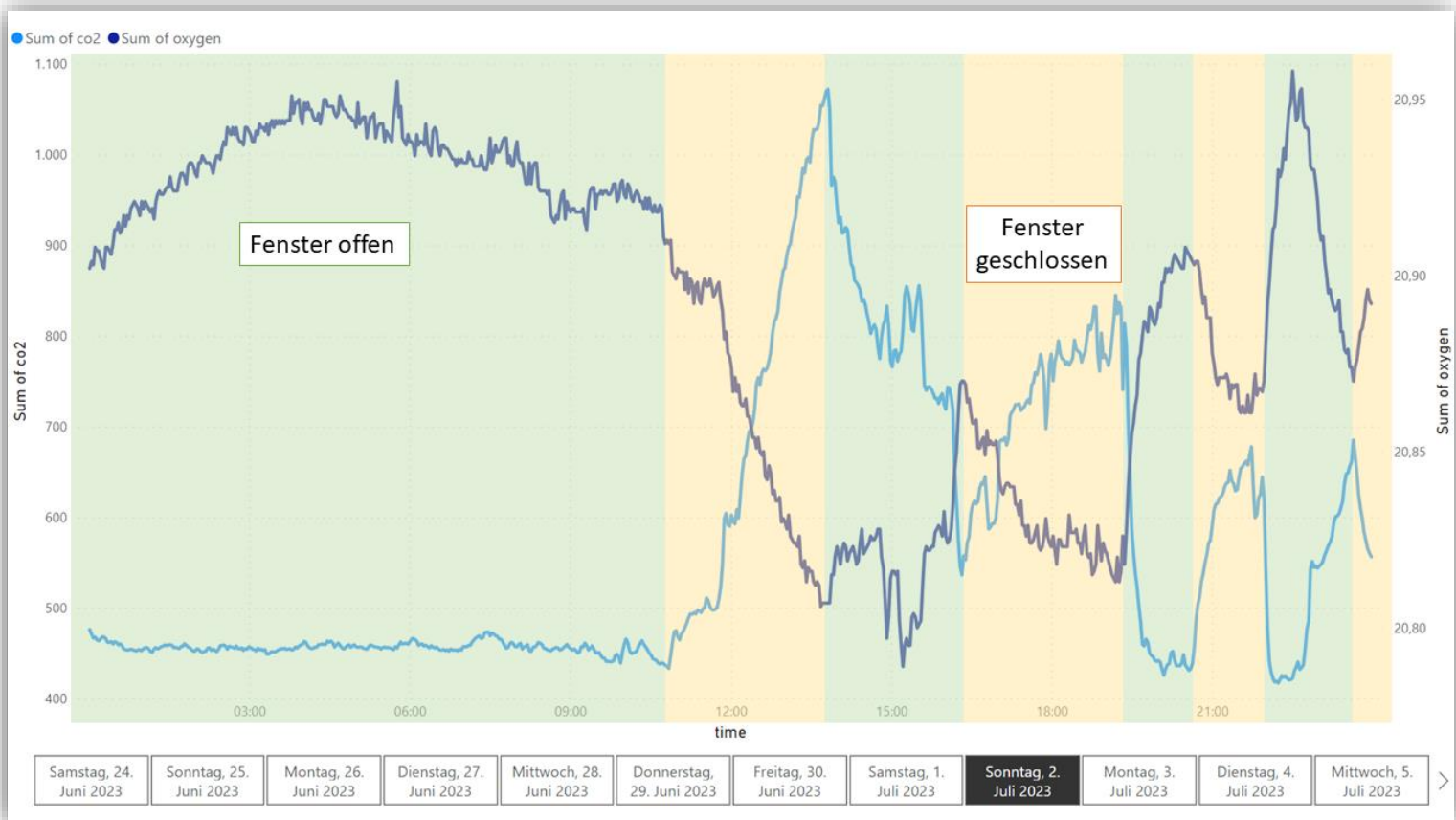
Wir nehmen an, dass sich das beobachtete Phänomen unter veränderten Rahmenbedingungen, wie zum Beispiel während der Wintermonate, wenn Fenster tendenziell geschlossen bleiben, anders darstellen könnte. Diese Bedingungen könnten potenziell dazu beitragen, die Auswirkungen des Kochens auf die Luftfeuchtigkeits- und VOC-Werte klarer hervortreten zu lassen. Zukünftige Untersuchungen könnten daher darauf abzielen, diesen Aspekt genauer zu beleuchten.



5.2.6 Lüftungsvorgänge lassen sich klar in den Daten erkennen

Im Rahmen unserer Analyse untersuchten wir den Zusammenhang zwischen dem CO₂-Gehalt und dem Sauerstoffgehalt, insbesondere während der Nachtstunden, wenn sich eine schlafende Person in der Wohnung befindet. Eine intuitive Vermutung war, dass mit einer Zunahme des CO₂-Gehalts ein entsprechender Rückgang des Sauerstoffgehalts beobachtet werden könnte. Allerdings konnten wir in unseren Untersuchungen dieses erwartete Muster nicht feststellen. Der Hauptgrund für diese Beobachtung liegt in der Praxis des ständigen Lüftens während der Nacht am Messort. Durch das offene Fenster konnte stets frische Luft in den Raum gelangen, was zu einer kontinuierlichen Durchmischung und somit zu einer relativ stabilen Sauerstoffkonzentration führte, unabhängig vom CO₂-Gehalt.

Trotzdem haben unsere Untersuchungen wertvolle Erkenntnisse geliefert. Insbesondere konnten wir feststellen, dass die analysierten Messwerte eine hohe Sensitivität gegenüber Lüftungsverhalten aufweisen. Es ist uns dadurch möglich, zweifelsfrei zu erkennen, wann ein Raum gelüftet wird und wann alle Fenster geschlossen sind. Diese Beobachtung kann sowohl für die Überwachung des Lüftungsverhaltens als auch für zukünftige Studien, die den Einfluss von Lüftungsverhalten auf andere Parameter untersuchen, von großem Nutzen sein.

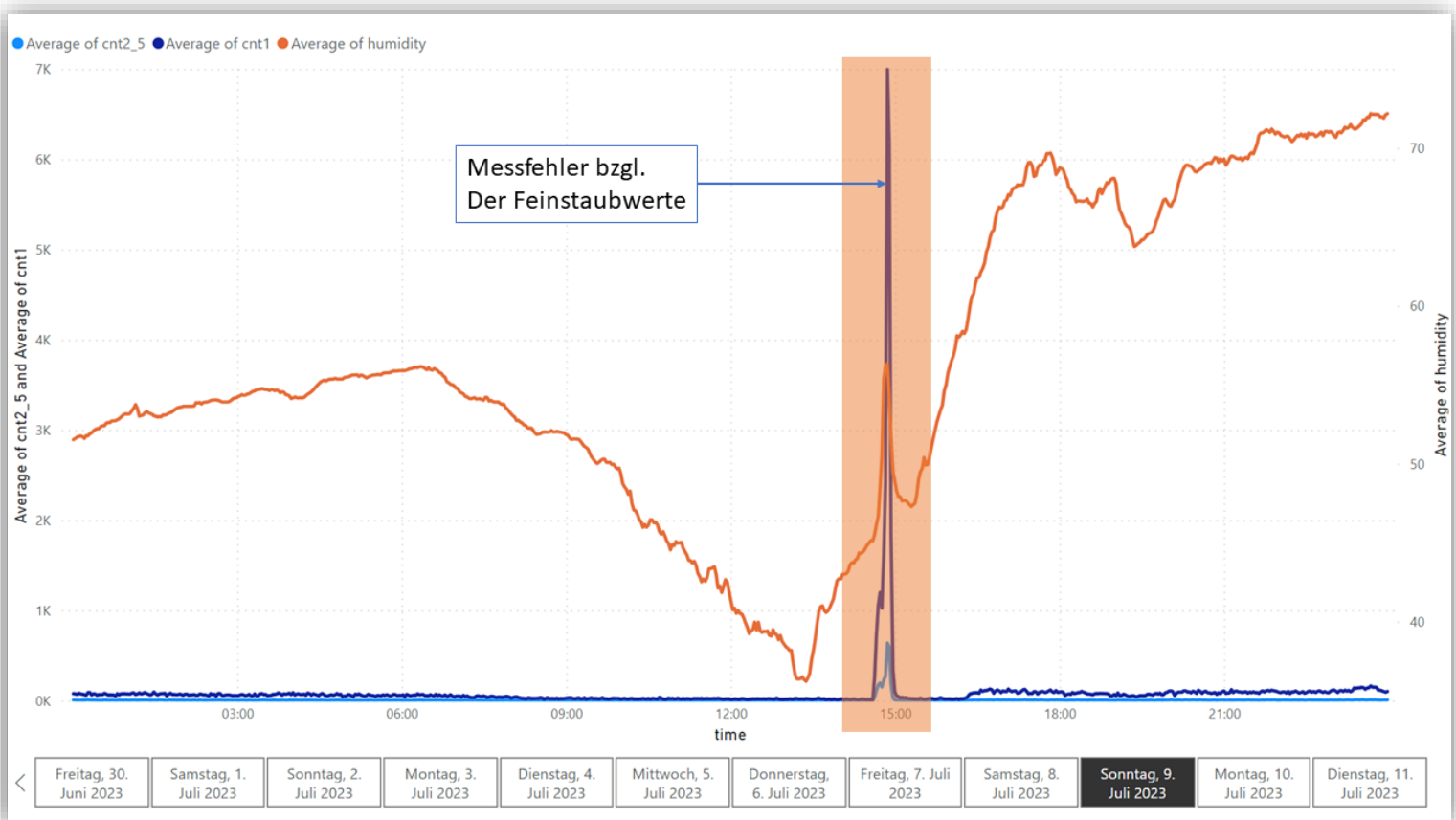


5.2.7 Messfehler bzgl. Feinstaubwerte ausgelöst durch Luftfeuchtigkeit

Am Sonntag, den 09.07.23 konnte ein Messfehler bzgl. der Feinstaubwerte beobachtet werden. Dieser Tag war ungewöhnlich heiß. Der plötzlich auftretende Starkregen, kurz vor 15:00 Uhr, hat daraufhin zu einem sprunghaften Anstieg der Luftfeuchtigkeit geführt.

Der Messfehler lässt sich dadurch erklären, dass bei Wasserdampf bzw. Feuchtigkeit in der Luft die Wassertröpfchen eine ungefähre Größe von 1-3 μm aufweisen. Genau diese Feinstaubpartikel sind auch sprunghaft angestiegen.

Interessant ist, dass der Kurvenverlauf darauf schließen lässt, dass ein Messfehler nur dann auftritt, wenn es einen stark sprunghaften Anstieg der Luftfeuchtigkeit gibt. Wie man im Bild erkennen kann, ist die Luftfeuchtigkeit nach dem ersten Peak noch einmal kontinuierlich angestiegen. Dies hat jedoch nicht zu einem weiteren Messfehler geführt. Wobei jedoch ab einer Luftfeuchtigkeit von 60% auch wieder die Zählung der Partikel der Größe von einem Mikrometer ansteigt.



5.3 Analyse des dritten Datensatzes

Im dritten Datensatz wurden zuletzt noch die folgenden Szenarien erhoben und miteinander verglichen:

- Eine Kochaktivität mit einem Cerankochfeld
- Ein Kochvorgang mit einem Gasherd

Wir hatten eigentlich eine deutliche Unterscheidung der Luftwerte zwischen dem Kochvorgang mit Gasherd und Cerankochfeld erwartet da momentan auch in den USA heiß diskutiert wird den Gasherd aufgrund von Verursachung schlechter Luftqualität in Innenräumen zu verbieten. Jedoch konnten wir eine Verschlechterung der Luftqualität nicht nachweisen.

Bei der Analyse wurden dabei die Metriken NO₂ (Stickstoffdioxid) und SO₂ (Schwefeldioxid) sowie die verschiedenen Werte der Feinstaubpartikel, betrachtet.

Möglicherweise gab es unterschiedliche Messbedingungen wie z.B. verschiedene Dunstabzugshauben oder geöffnete Fenster. Zudem kann auch entscheidend sein welches Gericht gekocht wird.

Um hier eine valide Datenbasis zu schaffen die einen Vergleich der Daten zulässt sollte also bei zukünftigen Messerhebungen auf die oben genannten Bedingungen eingegangen werden.

Ansonsten wurden auch in diesem Datensatz wieder vergleichbare Korrelationen und Zusammenhänge identifiziert, die in den Zwei Datensätzen zuvor schon ausführlich erläutert wurden.

6. Weiterführung und Ausbau des Projekts

Die in diesem Projekt erarbeitete Grundlage zur Messung und Analyse von Umwelt- und Partikelmetriken bietet eine Vielzahl an Möglichkeiten zur Vertiefung und Erweiterung. Im Folgenden skizzieren wir einige Möglichkeiten zur Weiterentwicklung und Ausweitung des Projekts.

6.1 Erweiterte Messungen

Die Erhebung zusätzlicher Daten bietet eine Fülle an Möglichkeiten zur Untersuchung der Luftqualität unter verschiedenen Bedingungen und während unterschiedlicher Aktivitäten. Diese könnten das Kochen mit verschiedenen Methoden, das Lüften, die Verwendung von Haushaltsprodukten, menschliche Anwesenheit und Bewegung im Raum und vieles mehr umfassen. Durch die Sammlung und Analyse solcher Daten könnten wir ein detaillierteres Bild der Faktoren zeichnen, die die Innenraumluftqualität beeinflussen und ggf. neue Korrelationen und Messfehler aufdecken.

6.2 Implementierung eines Analyse-Dashboards

Ein Analyse-Dashboard, etwa implementiert mit Power BI, könnte eine effiziente und benutzerfreundliche Möglichkeit sein, die erhobenen Daten zu visualisieren und zu analysieren. Ein interaktives Dashboard ermöglicht es den Nutzern, die Daten nach ihren individuellen Anforderungen zu filtern und zu untersuchen. Damit können sie Erkenntnisse gewinnen, die spezifisch auf ihre Interessen und Bedürfnisse zugeschnitten sind. Ein Dashboard würde auch die Hürde der momentanen Analyse und Bedienung per Code-Funktionen eliminieren und eine intuitive Datenanalyse ermöglichen.

Dazu müsste die bestehende Datenbank in Power BI eingebunden werden, ein geeignetes Datenmodell in Power BI geriert werden und daraufhin ein intuitiver Report bzw. Dashboard auf Basis der Nutzeranforderungen geplant und erstellt werden.

6.3 Automatisierte Datenanalyse und Erweiterung der Software-Funktionen

Das derzeitige Programm könnte um Funktionen für eine automatisierte Datenanalyse erweitert werden. Eine implementierte Data Analysis Pipeline könnte dabei helfen, Daten automatisch zu importieren, aufzubereiten, zu analysieren und Berichte zu generieren. Das würde den Analyseprozess effizienter gestalten und könnte dazu beitragen, noch tiefere und vielseitigere Erkenntnisse aus den Daten zu gewinnen. Dieser Ansatz ist gängig in sämtlichen Data Science Projekten. Der Grundstein dafür ist durch die von uns entwickelten Klassen schon vorhanden.

6.4 Anwendung von Machine Learning

Die Anwendung von Machine Learning-Techniken könnte einen bedeutenden Schritt in der Weiterentwicklung des Projekts darstellen. Durch den Einsatz von Entscheidungsbäumen könnte man beispielsweise versuchen, die Metriken "Performance" und "Health" zu rekonstruieren oder zu verbessern. Darüber hinaus könnte man versuchen, spezifische Vorgänge oder Muster in den Daten zu identifizieren, etwa das Lüften, Kochen, das Vorhandensein von Messfehlern oder die Anwesenheit von Personen im Raum.

Um dies umzusetzen, gibt es zwei Herangehensweisen.

1. Man müsste neue Messungen durchführen und im Nachhinein den Datensatz mit Labels der jeweiligen durchgeführten Aktivität oder Messfehler, in diesem Zeitraum, versehen. Dies ermöglicht einen Machine Learning Ansatz mit Algorithmen des Supervised Learning. Für diesen Anwendungsfall wären das Klassifikationsalgorithmen. Dabei bietet sich vermutlich ein Entscheidungsbaum oder ein Neuronales Netz an, da die Metriken einen Grenzwert besitzen, ab dem der Wert als kritisch betrachtet wird. Die Labels sind notwendig, damit sich der Algorithmus selbstständig korrigieren und damit auch verbessern kann.
2. Ein weiteres Vorgehen ist dem Unsupervised Learning zuzuordnen. Hierbei werden die Daten zuvor nicht gelabelt und es werden Machine Learning Ansätze herangezogen, die selbstständig Muster und Cluster in den Daten identifizieren.

Es ist auch ein kombinierter Ansatz aus beiden Lernformen möglich.

6.5 Reengineering des airQ Sensors

Eine weitere Möglichkeit wäre das komplette Reengineering des Sensorsystems selbst. Durch den Einsatz von z.B. Raspberry Pi und externen Sensoren könnte der airQ Sensor nachgebildet werden. Dies erfordert jedoch ein gutes Verständnis über Sensorik und Elektrotechnik. Jedoch hat dieses Projekt sehr viel Potential im Wissenserwerb über Sensorik und das Planen von neuen Systemen.

7. Fazit

Im Verlauf dieses Analyseprojekts wurde eine fundierte und umfangreiche Untersuchung der Luftqualität unter verschiedenen Bedingungen und in unterschiedlichen Kontexten durchgeführt. Durch die Verwendung hochwertiger Sensortechnologie und der Entwicklung einer leistungsstarken Schnittstelle konnte eine Vielzahl von Metriken erfasst und analysiert werden.

Es wurden Module programmiert, die es dem Nutzer erlaubt auch ohne großes Vorwissen, selbstständig eine eigene Datenanalyse vorzunehmen. Zudem wurden die Module in vollem Umfang und nach einem hohen Standard dokumentiert, um eine unkomplizierte Weiterentwicklung zu ermöglichen.

Die Untersuchung hat ergeben, dass es signifikante Korrelationen zwischen verschiedenen Umweltfaktoren und Luftpartikeln gibt, die sich sowohl im Labor- als auch im Einpersonenhaushaltskontext widerspiegeln. Dazu zählen z.B. Zusammenhänge zwischen Feinstaubpartikeln, Temperatur, Feuchtigkeit und Taupunkt. Diese Erkenntnisse eröffnen neue Möglichkeiten zur Verbesserung der Raumluftqualität und tragen dazu bei, das Bewusstsein für die Auswirkungen unserer Aktivitäten und Umweltbedingungen auf die Innenraumluft zu schärfen.

Ebenso interessant war die Erkennung spezifischer Ereignisse durch die Analyse der Messdaten. Ein Stromausfall, ein Großbrand und sogar Lüftungsvorgänge konnten durch die aufgezeichneten Daten nachvollzogen werden. Dies unterstreicht die Genauigkeit der verwendeten Sensorik und die Fähigkeit der eingesetzten Analysemethoden, relevante Ereignisse zu identifizieren.

Das Projekt hat jedoch auch Herausforderungen aufgezeigt, insbesondere in Bezug auf die Schwierigkeiten, bestimmte Aktivitäten wie Kochen eindeutig in den Daten abzugrenzen. Zudem sind Messfehler nicht einfach zu reproduzieren, um dadurch weitere Erkenntnisse zu gewinnen.

Die Weiterführung und der Ausbau des Projekts, beispielsweise durch die Implementierung eines Analyse-Dashboards, die Automatisierung der Datenanalyse und die Anwendung von Machine Learning, bietet spannende Möglichkeiten, die Qualität und Tiefe der Analyse in Zukunft weiter zu verbessern.

Dieses Projekt hat uns dazu veranlasst, unsere Fähigkeiten und unser Verständnis auf dem Gebiet der Datenanalyse sowie im speziellen Bereich der Luftqualitätsüberwachung erheblich zu erweitern. Es hat uns gezeigt, wie die Verbindung von technischem Fortschritt und analytischer Präzision zu bedeutsamen Erkenntnissen führen kann, die in praktischen Anwendungen einen echten Unterschied machen können.

Die Möglichkeit, diese anspruchsvolle Aufgabe zu bewältigen und uns in diesen hochinteressanten und relevanten Forschungsbereichen zu beweisen, war eine wertvolle Erfahrung. Diese Erkenntnisse werden uns bei unseren zukünftigen Projekten und unserer weiteren Entwicklung zugutekommen. Wir möchten diese Gelegenheit nutzen, um uns bei Frau Gligorevic und der FH Aachen für die uns gebotene Chance zu bedanken.