Phenomenon# 9 مشاہدہ #۹

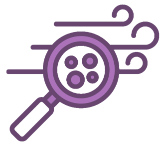
The Air We Breath

ہماری سانسوں کا معیار

Muhammad Bilal and Muhammad Sabieh Anwar

June 10, 2025

This manual is your guide on your journey from understanding the concept of air quality, to collecting data, and finally to interpreting and acting on your findings. Moving step-by-step through the sections, you'll find explanations, reflection questions, and tools to help you engage with real data, just like a scientist. Here is a map of your journey.



When you are traveling through your daily route, you can observe and sense everything around you. You observe the bustling flow of traffic. You feel the cool breeze and the warmth of the sun. But what about the things we cannot see or sense? Things around us that are tiny, almost invisible. Have you ever thought about what you're directly breathing in?

What do you think makes up the air around us?

We have read in our books that air is made up of oxygen, nitrogen, carbon dioxide along with some other gases. There are also natural solid particles such as bacteria, pollen and mineral dust. But we are missing something. This is the stuff we have added in the air ourselves, through our fireplaces, cooking stoves, factories and vehicles releasing smoke of different types into the environment. Today, we are going to take a closer look at the air around us and understand the changes human activity has brought about.

1. Particulate Matter (PM)

Our air is filled with tiny and often harmful particles called particulate matter (PM). Lahore and Punjab encounter a hazardous smog every year, largely caused by these particles.

Many of the particles are too small to be seen with the naked eye but still pose a serious threat to human health. It is reported that PM has a negative impact on life expectancy.[ref]

These particles are generally classified into three different groups:

- PM 1.0
- PM 2.5
- PM 10.0

The number after PM refers to the *maximum* diameter of the particles in micrometers.

To get an idea of how small these particles are, think about this. A single PM 2.5 particle is about 20 times smaller than the average width of a human hair. No wonder we can't see them!

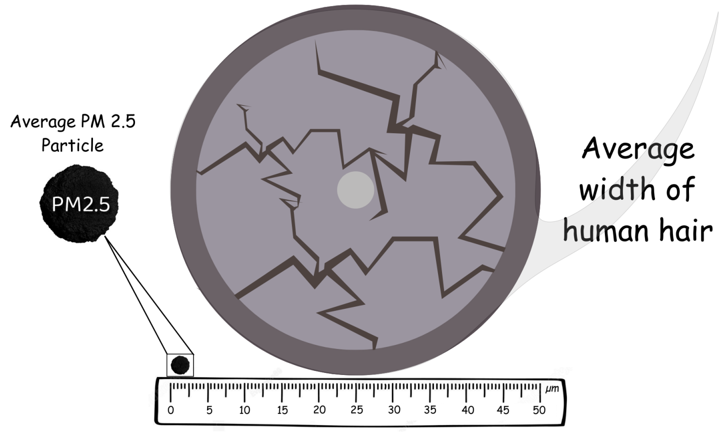


Figure 1: Size of an average PM 2.5 particle.

These particles are the main culprits behind our hazardous air quality and its harmful byproducts such as the smog we see in winters. Among these, PM2.5 is the most concerning one for us as humans because of its small size. Thankfully, the methods needed to measure and analyze PM2.5 concentration and study their harmful effects are well understood now. To study more, you can visitEnvironmental Protection Agency's website.
These particles did not reach dangerously high levels on their own. On our journey towards a better future, we have taken paths that have forced both nature and ourselves to pay the

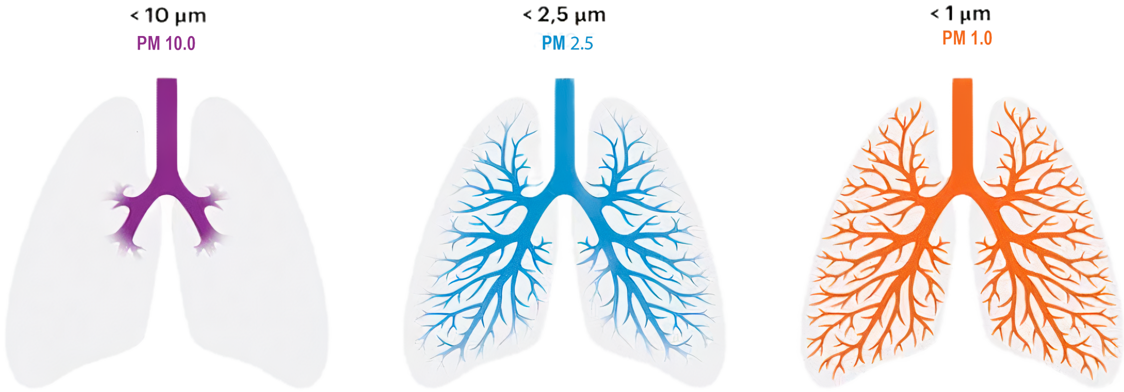


Figure 2:
An illustrative approach to better visualize that the smaller the particles are, the deeper they go into the lungs [1].

price. Now, more than ever, we must choose cleaner, more sustainable ways to move forward. Our religion also places great emphasis on cleanliness and a clean environment, considering it as fulfilling half of one's faith.

The Prophet (peace be upon him) paid special emphasis to cleanliness. By extension, we need to ensure that activities do not add hazard to the environment, including polluting the air we breathe.

الطُّهُورُ شَطْرُ الإِيمَانِ

صفائی نصف ایمان ہے۔
- الحدیث

Cleanliness is half of faith.
- Al Hadith

Before we dive further into this, let's first visualize what is actually going on in the air around us.

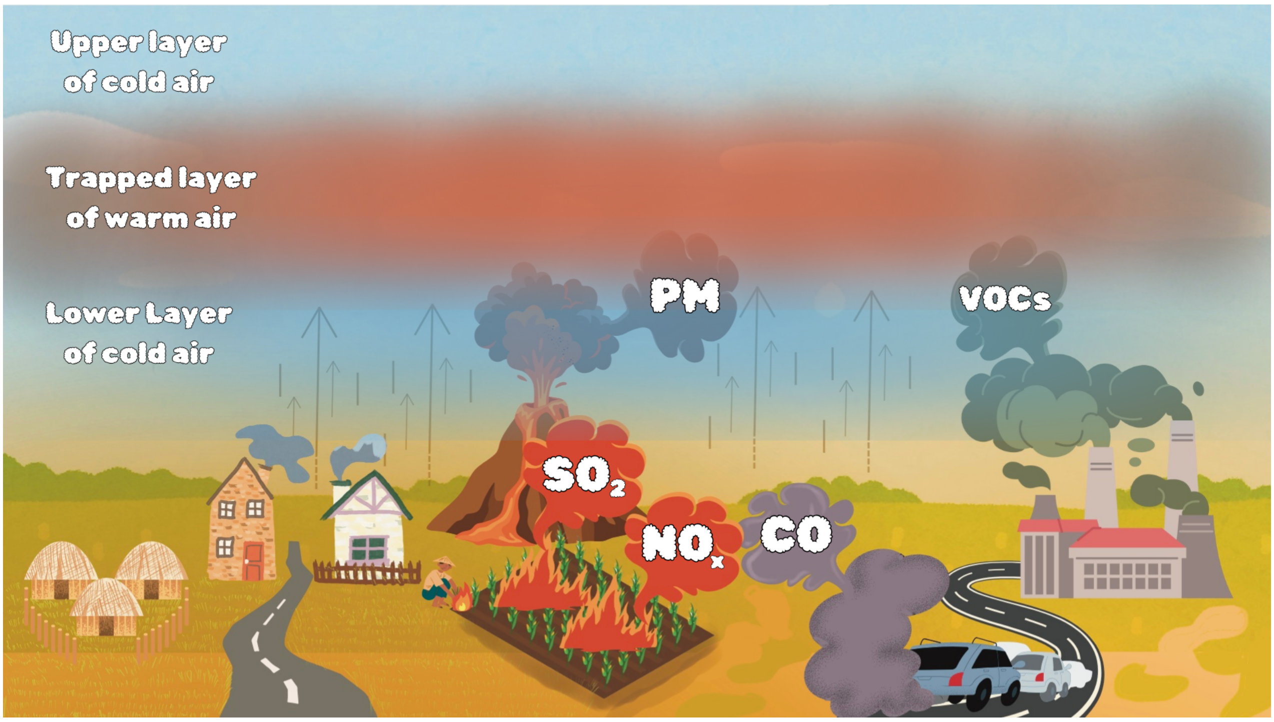


Figure 3: In this figure, we see various sources of air pollution such as vehicular emissions, industrial smoke as well as natural activity like volcanic eruptions—releasing harmful pollutants such as particulate matter (PM), sulfur dioxide (SO_2), nitrogen oxides (NO_X), carbon monoxide (CO), and volatile organic compounds (VOCs). VOCs are a group of chemicals that easily evaporate into the air and contribute to the formation of smog. An example of a VOC is the smell from fresh paint or gasoline, which releases these gases into the air. These pollutants get trapped near the ground during a condition known as temperature inversion, where a layer of warm air becomes trapped between two layers of cooler air. This warm layer acts as a lid, preventing the dispersion of pollutants. As a result, contaminants remain confined in the lower atmosphere, leading to poor air quality.

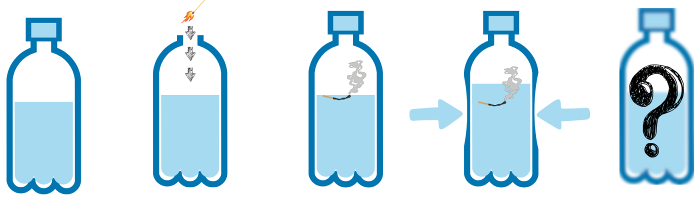
[Q 1.] You might have noticed that smog only appears during winters but not in other seasons. Can you deduce the reason behind it from the *figure 3*?

[Q 2.] If the season were warmer, how would the environment change, and how would this impact smog formation?

ایک دلچسپ مشاہدہ

ایک خالی بوتل لیں اور اس کا ۴/۳ حصہ پانی سے بھر دیں۔ ایک ماچس جلائیں اور اسے بوتل کے اندر پھینک دیں۔

فوراً بوتل کو اچھی طرح بند کر دیں۔ اب بوتل کو دبائیں اور پھر چھوڑ دیں۔ کیا بوتل کے اندر کوئی چیز بدلی؟



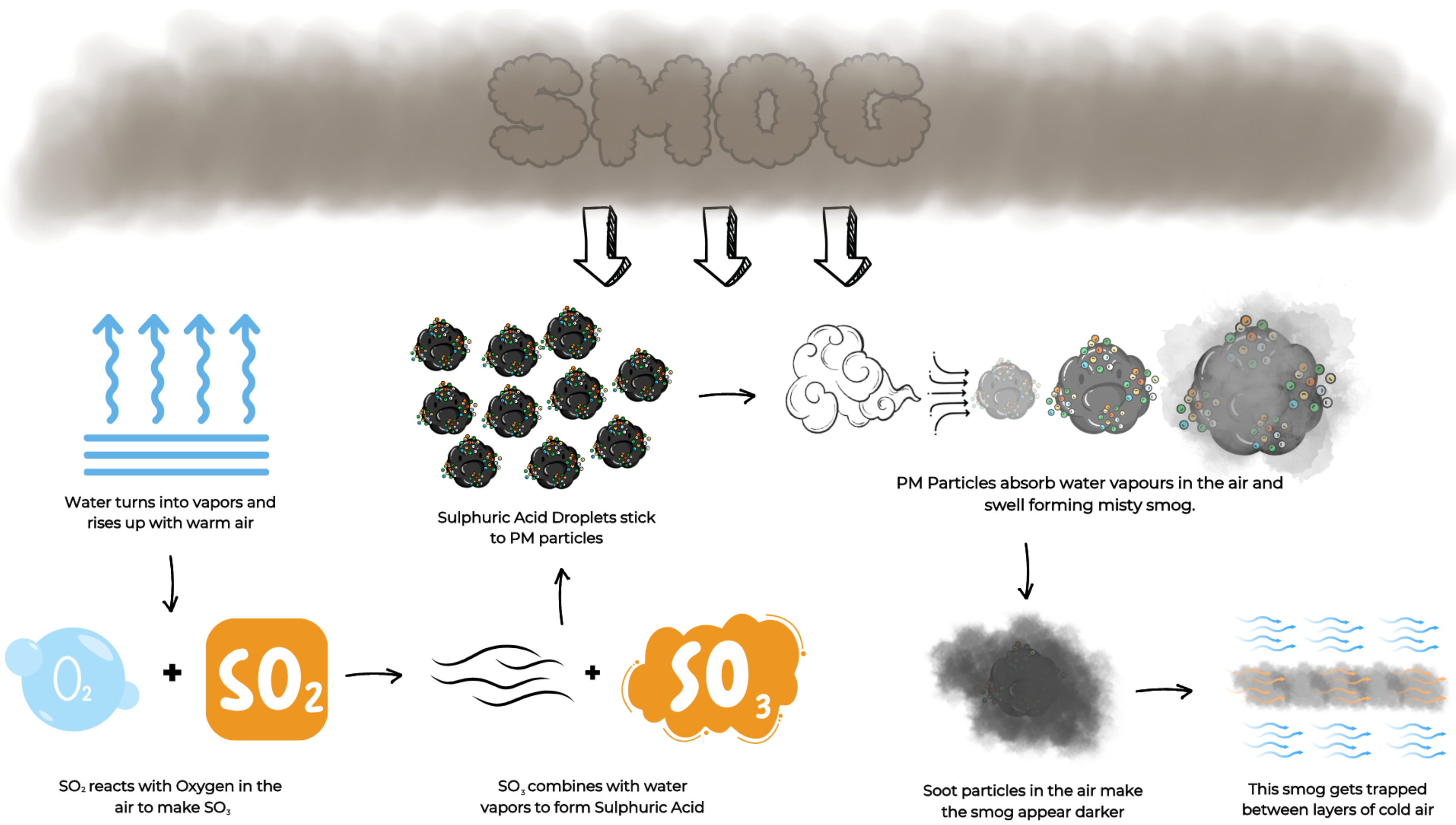


Figure 4: Smog forms below the cold layer of air when trapped pollutants—such as sulfur compounds, VOCs, and particulate matter—combine with water vapor in the atmosphere, creating misty, dark layers of pollution that settle near ground level

To analyze the quality of our air quantitatively, scientists measure the concentration of PM particles and calculate a number called the Air Quality Index (AQI). Today, as aspiring scientists we are going to do just that using a PM Tracker developed by Khwarizmi Science Society and Qosain Scientific. Let's uncover the secrets of the air we breathe!

2. What we need

Since digital devices have made our lives much easier, we only need two things to analyze our environment:

1. The PM Tracker—our scientific tool to measure PM concentrations (head to Qosain Scientific's website [2] to access the setup guide for the tracker).

2. A laptop or computer with MS Excel or your favourite data processing software — to analyze the collected data and calculate the air quality index (AQI), measure uncertainties and analyze real-life scenarios that generate hazardous particulate matter.

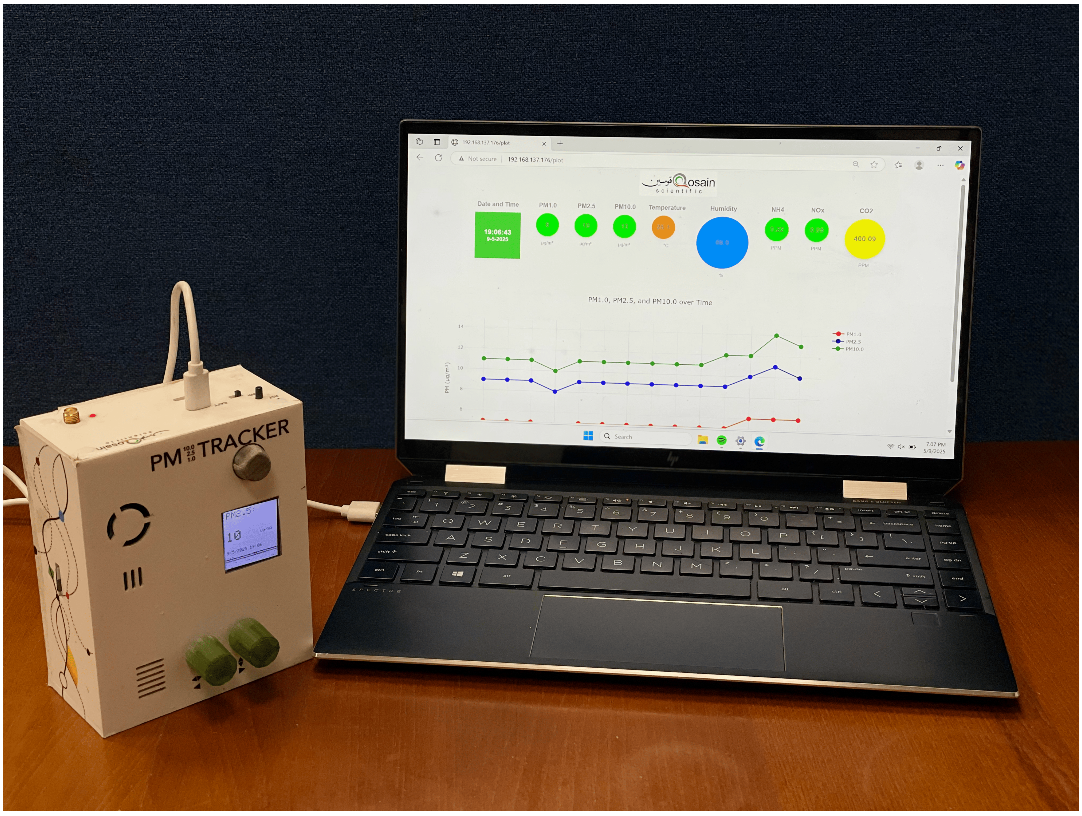


Figure 5: Real-time plot of a PM tracker connected to a laptop via local wireless/WiFi network connection.

3. Collecting data

1. Place your PM Tracker in an area whose air quality you wish to measure. Maybe it is near a window or in a classroom after break time.

2. Set up the PM Tracker and start recording.

3. Let it gather data for some time (the duration for which the tracker collects data is entirely up to you).

4. Stop the measurement and download the data file. The file will have a name that includes the date and time of the recording.

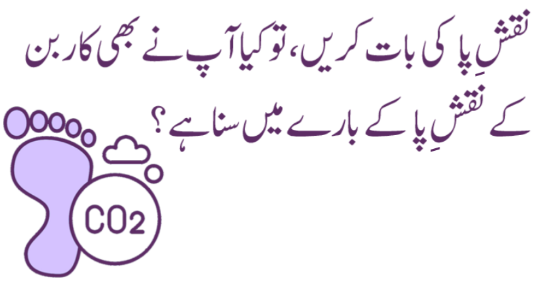
If you open your dataset in a numerical processing software, such as Excel, the data may look like this:

ms	UTC_timestamp	temperature	humidity	PM1.0	PM2.5	PM10.0
0	2025-1-12T00:00:00+00:00	20.2	75	68	118	149
6395	2025-1-12T00:00:07+00:00	20.5	73	74	115	153
8912	2025-1-12T00:00:09+00:00	20.6	73	62	111	138
16397	2025-1-12T00:00:17+00:00	20.6	73	62	100	131
18919	2025-1-12T00:00:19+00:00	20.7	73	64	99	132
26399	2025-1-12T00:00:26+00:00	20.5	73	56	93	120
28907	2025-1-12T00:00:29+00:00	20.7	73	55	99	123
36412	2025-1-12T00:00:36+00:00	20.7	73	58	103	128
38916	2025-1-12T00:00:39+00:00	20.8	72	61	104	132
46420	2025-1-12T00:00:47+00:00	20.8	73	62	99	129
48927	2025-1-12T00:00:49+00:00	20.8	72	61	98	128
56423	2025-1-12T00:00:56+00:00	20.9	72	61	98	128
58928	2025-1-12T00:00:59+00:00	21	73	56	88	116
66431	2025-1-12T00:01:06+00:00	21	73	59	93	123
68940	2025-1-12T00:01:09+00:00	21	74	61	95	126
76435	2025-1-12T00:01:17+00:00	21.1	74	55	94	119
78948	2025-1-12T00:01:19+00:00	21	74	60	93	123

Figure 6: The figure shows sample readings collected from the PM tracker.

How do we extract meaningful results from such a large dataset?

If you look at the data, there are multiple measurements after regular intervals. Each measurement is like a footprint, giving us clues about what's happening in the air. By analyzing this data with our really interesting statistical tools, we can uncover patterns and create or develop new insights. Let's look at the tools at our disposal.



4. Statistical Tools

4.1. Average:

The average value tells us the typical value of our data set. To understand it better, you can see how a player's performance in cricket can vary from match to match. Sometimes, they score a lot of runs, while in other matches, they might get out early with only a few. A player's runs are not always the same. If we want to understand how well a player performs overall, we look at their average score. The averages provides a better idea of his consistency and skill rather than just looking at one or two games.

To find the average of your data set, you have to add all the values and divide by the number of values you have:

$$\textbf{Average } (\mu) = \frac{\textbf{Sum of all observations}}{\textbf{Total number of observations}}$$

In Excel:

- To calculate the average value of your dataset using Excel, you can take help of the built-in average function. Type `=Average()` in an empty cell.
- Then place the cursor between the brackets and select the data you want to calculate the average for, by first clicking on the first value and then holding `"Shift"` key and clicking on the last value that you want.
- This will define the range for your formula which will automatically calculate the average value of the selected dataset for you.

temperatu	humidity	PM1.0	PM2.5	PM10.0		PM2.5 Avg
20.9	58	172	305	381		=AVERAGE(I2:I386)
20.8	59	185	306	394		
20.9	58	172	284	367		

Note:

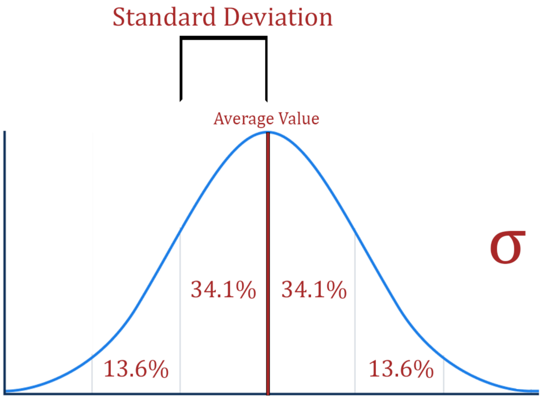
When performing such calculations as average, remember that your final answer should only be as precise as your original measurements. This makes sense, right? If the smallest reading your instrument can measure with an instrument, say a measuring scale is 1 cm, no calculation can give a more precise answer. To visualize this in a different way, imagine you're using a smartphone camera that can only capture images at a certain resolution. No matter how many pictures you take or how you combine them, you can't see details beyond that resolution. So, if your instrument measures values precisely up to one decimal place, then even if your calculations give you extra digits, you should round the result to match the precision of the instrument. This way, your readings reflect exactly what the measurement tool can truly tell you.

[**Q 3.**]What does a large (or small) average PM2.5 value tell you about the air quality?

4.2. Standard deviation:

Standard deviation builds on the concept of average. It tells us how much variation exists in a set of data from the average value. If all the values are close to each other, the standard deviation is low. If the values are far apart, the standard deviation is large.

To calculate standard deviation, we use a formula that goes as:



$$\sigma = \sqrt{\frac{\sum |x-\mu|^2}{N}}$$

where:

- σ represents standard deviation,
- $\sum$ means "sum of",
- x is a value in the data set,
- μ is the average of the data set.
- and N is the number of data points in the population.

In Excel:

- To calculate the standard deviation of your dataset using Excel, you can use the built-in function. Type `=STDEV.P()` in an empty cell.
- Then place the cursor between the brackets and select the data you want to calculate the standard deviation of by first clicking on the first value and then holding `"Shift"` key and clicking on the last value that you want.
- This will define the range for your formula which will automatically calculate the average value of the selected dataset for you.

temperatu	humidity	PM1.0	PM2.5	PM10.0		Std. Dev PM 2.5
20.9	58	172	305	381		=STDEV.P(I2:I386)
20.8	59	185	306	394		
20.9	58	172	284	367		
21	59	157	281	350		
20.8	58	159	283	352		

[**Q** 4.] If your standard deviation value is very large or small, what does that tell you about your environment? What does it tell you about your data?

4.3. Noise in Data:

While analyzing our measurements, it's important to understand that not every variation we observe is meaningful. Some fluctuations happen simply due to noise—random changes caused by minor disturbances like vibrations or even small, momentary gusts of air. Recognizing noise helps us better interpret our results and avoid overthinking every small spike or dip in the data. Statistical tools like standard deviation and rolling averages help reduce the impact of this noise and highlight real trends.

5. Calculating AQI

With the average value of your data, you can now calculate the Air Quality Index (AQI) of your environment. In a standardized method of measurement, the AQI is an indicator of the air quality around us. We can calculate it using the AQI defined by EPA(Environmental Protection Agency) as follows:

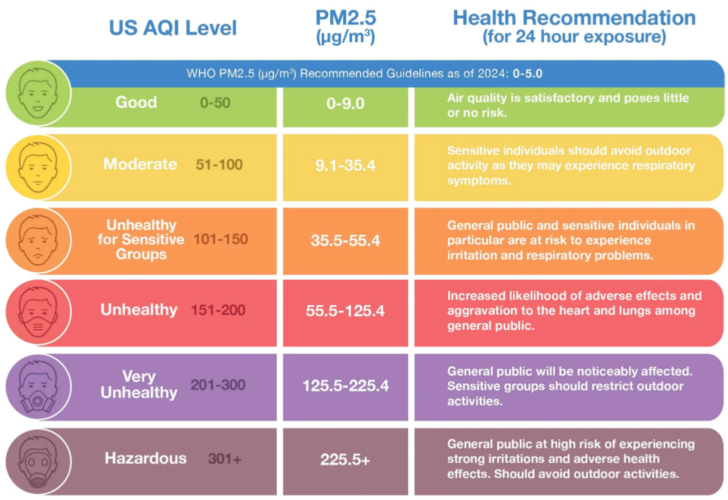


Figure 7: Air quality index chart with corresponding PM2.5 g/m^3 outlining U.S. AQI levels for PM2.5, showing pollution breakpoints((the concentrations corresponding to the upper and lower limits of the different AQI levels)) and associated health recommendations for 24-hour exposure. As PM2.5 levels rise, air quality worsens—ranging from "Good" (0–9.0) to "Hazardous" (225.5+). Each level indicates increasing health risks, especially for sensitive groups, with protective guidance based on pollutant concentration [3].

The formula for calculating AQI is:

$$AQI = \frac{(A_h - A_l)}{(BP_h - BP_l)} \times (C_p - BP_l) + A_l$$

where,

C_p = the concentration of PM2.5
BP_h = the higher PM2.5 concentration break point corresponding to C_p
BP_l = the lower PM2.5 concentration break point corresponding to C_p
A_h = the higher AQI value corresponding to BP_h
A_l = the lower AQI value corresponding to BP_l

Example:

Suppose the average value of PM2.5 concentration is 167 μ g/m^3 which corresponds to the AQI range 201–300. This means that $A_l = 125.5$, $A_h = 225.4$, $BP_l = 201$, $BP_h = 300$. With these values, our formula yields:

$$AQI = \frac{(300 - 201)}{(225.4 - 125.5)} \times (167 - 125.5) + 201 = 242.13$$

In this way, you will be able to calculate AQI for your surroundings from PM2.5 concentration. Note that the AQI value does not have any units. This is because AQI is a standardized scale that represents air pollution levels, not the concentration of pollutants themselves.

6. Some advanced tools

Average and standard deviation alone can help you calculate AQI and develop insights, but it does not hurt to know more. After all, there is so much more hidden in the data we have collected with our PM trackers.

6.1. Rolling average

Rolling average is like a moving window. Imagine you are looking our of the window of your car as the car is moving. At any moment, you can only see the same amount of scenery but that scenery is constantly changing with time. At any moment, you lose a bit of the previous scene as you get more information of what's ahead. Similarly, when we are calculating rolling average, we drop the oldest number every time we get a new number and calculate the average again. Using this, we can see if our value is really increasing or just jumping around! It helps us differentiate between long lived or real effects from temporary or fleeting effects.

We obtain PM values every, say 2.5 seconds. This could be too fast and the PM concentration changes may be more gradual. So, instead of looking at just one number, we take the average of the last few numbers—maybe the last 1 minute or 5 minutes or maybe an hour. This way, we smooth out the ups and downs, perhaps averaging out noise. We will illustrate this averaging effect on real data in figure 7 and see the overall trend.

6.2. Plotting the data:

The AQI informs us about the general air quality of our surroundings. But that is just based on a single number, which is the average of the entire data set. How does the data change? If we want to analyze the trend throughout the period of data collection and deduce the factors behind poor air quality, we need to plot our data. This visualization gives new insights and exposes behavior that may go unnoticed when merely looking at the numbers. This visualization gives new insights and exposes behavior that may go unnoticed when merely looking at the numbers.

For that, we will use the "Plot" feature of Excel.

- First select the data range along with the corresponding timestamps you want to plot.
- You can get the time from UTC Timestamp by using the formula `=MID(B2,12,8)`.
- Once selected, press the `F11` key on your keyboard.
- Excel will plot your data on your screen. Click on `Change Chart Type` on the top bar. Select `Line` from the list and press ok.
- This chart will help you analyze the trend of air quality along with any unusual spikes or dips that might have occurred. You can deduce the reason behind those spikes or dips by analyzing the environment. By pressing the `+` sign on the top right corner, you can add error bars like standard deviation, rolling average, standard error in mean etc. from the `Error bars` sub-menu.

Your graph should look something like figure 8.

[**Q** 5.] What does the rolling average trend line show us in the graph?

[**Q** 6.] What happens to the behavior of a rolling average in the graph when you change its period length?

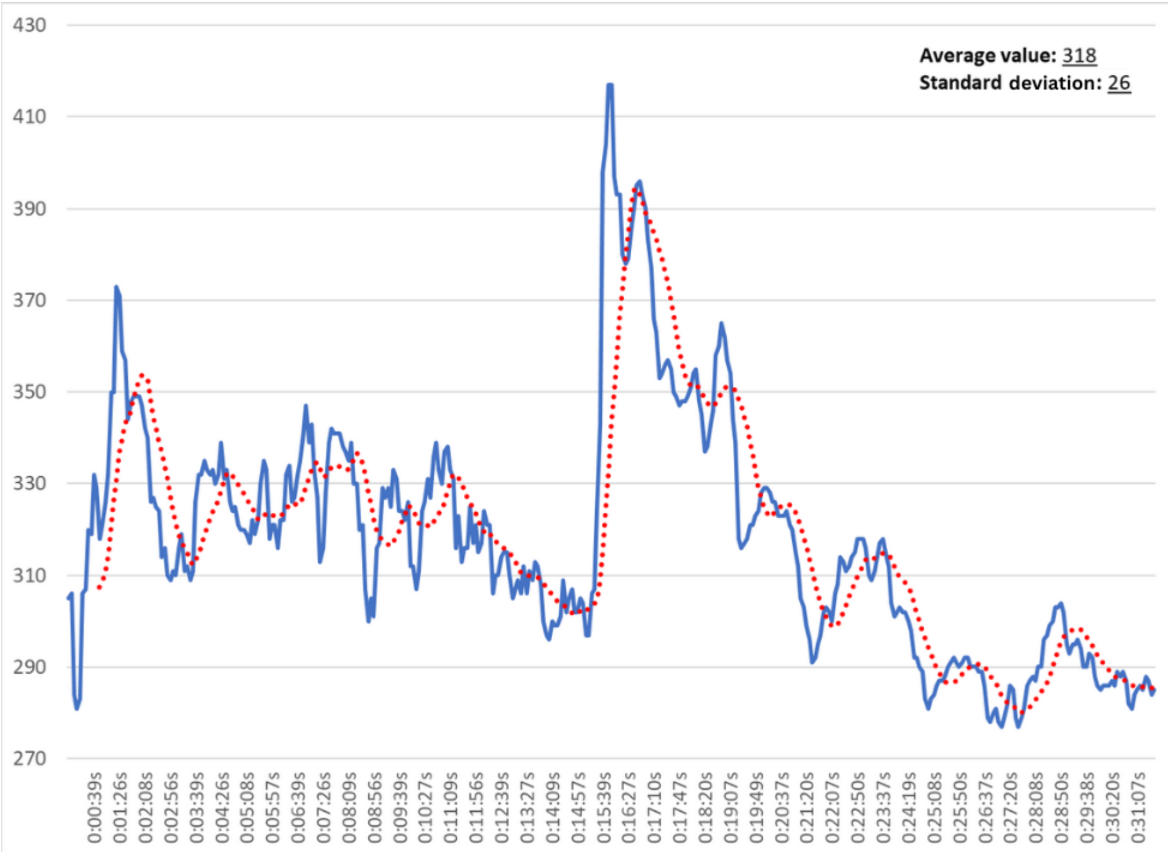


Figure 8: Plot of PM2.5 concentration (blue line) with rolling average (red dotted line) over time.The blue line shows the raw data, which includes both real changes in air quality and random noise. The red dotted line, representing the rolling average, smooths out short-term fluctuations and highlights the overall trend in the data, making it easier to interpret meaningful patterns.

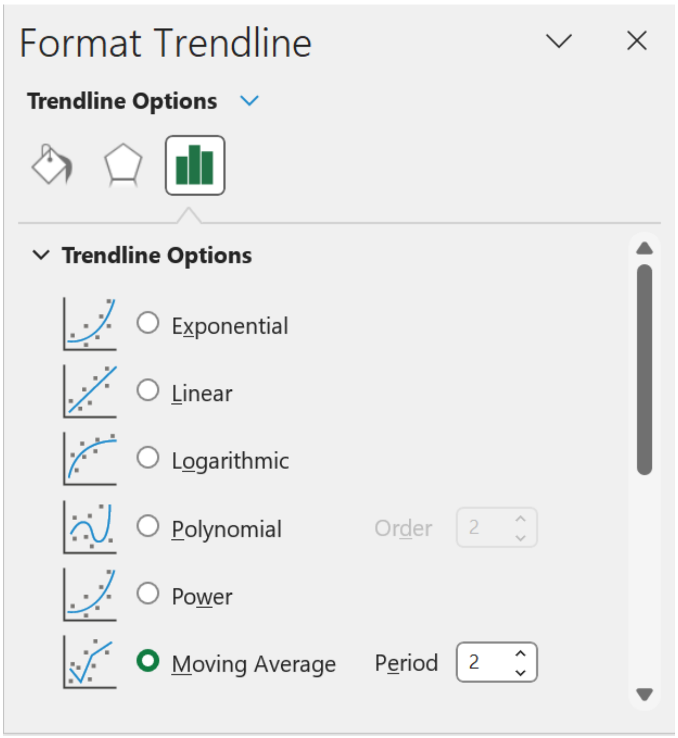


Figure 9: Adjusting period length of rolling average using trendline options in Excel.

6.3. Histogram:

You can make a histogram of the dataset as well. A histogram is just another way of graphically represent your data. As each method has its own perks, this one does as well. You can make a histogram in the same way you plotted your PM2.5 concentration. Just select your data, go to **`Insert`**, then **`Charts`** and select **`Histogram`** from the menu. It should look something like what is shown in figure 9:

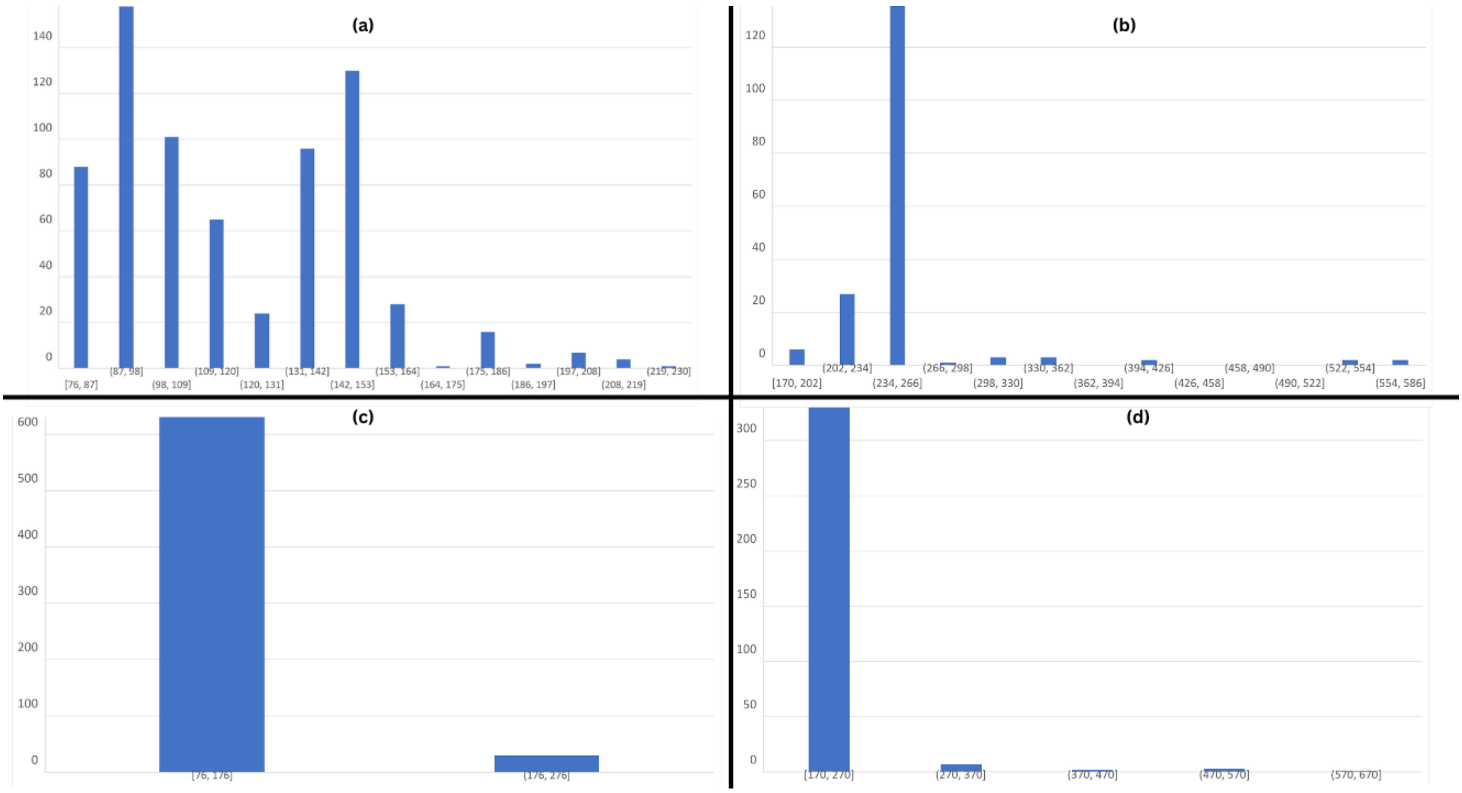
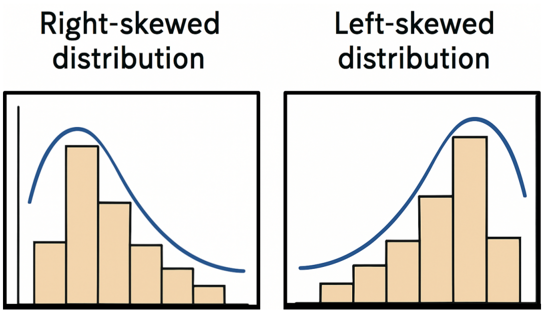


Figure 10: Each plot shows how the same type of data can look different depending on its natural spread and the bin width used. (a) Dataset with large spread and small bin width(17). (b) Dataset with small spread and small bin width(17). (c) Dataset with large spread and large bin width(100). (d) Dataset with small spread and large bin width(100).

A histogram can help you in quickly analyzing the following factors in your environment:

- **Spread**: A histogram helps you visualize how much your data is spread out. A lower spread means that the values are close to each other; a higher spread shows the opposite. Understanding both the average and the spread gives a fuller picture: two datasets can have the same average but very different spreads. A small spread means most values are close to the average, while a large spread indicates that values are scattered widely around the average.
- **Outliers**: You can also visualize if there were any potential extreme values in your dataset representing some extreme event.
- **Shape of distribution:** A histogram can have different shapes. For example, a normal (bell-shaped) distribution suggests that the values fluctuate around a stable mean, indicating a relatively controlled environment. A right-skewed distribution (long tail on the right) means most values are small, but a few large values cause the tail to stretch out. A left-skewed distribution (long tail on the left) means most values are large, but some occasional low values pull the tail to the left.

[Q 7.] What does the shape of the histogram tell you about your dataset? Can you deduce anything about the environment from its shape?



7. Bias in Data

Sometimes, the measurements we collect can consistently lean higher or lower than the true values. This systematic error is called bias. Bias is different from random noise—it pushes all measurements in a particular direction rather than just fluctuating randomly.

In a histogram, bias might show up as the whole distribution being shifted toward higher or lower values compared to what is actually happening in the environment.

For example:

If your PM Tracker's sensor is dusty or partially blocked, it might consistently report lower particulate readings than the true environmental concentration. Similarly, if the tracker is placed too close to a source of smoke (like near a kitchen exhaust) without reflecting the general room air, the readings could be biased high.

Recognizing potential sources of bias is important before interpreting results or comparing two datasets, so that the conclusions we draw are based on real environmental changes—not on hidden errors in measurement.

8. Correlation

If you have more than one PM trackers monitoring the same environment, you can get a deeper understanding of your environment as well as the accuracy of the sensors, by using another statistical tool called correlation. Technically correlation means that there is a relationship between two variables.

In statistics, correlation is the measure that expresses the extent to which two variables are related (meaning they change together). With a setup of at least two PM trackers, you can compare their data in excel by using the command `=Correl()` and inserting your data between the brackets. This will give you a single number answer. The closer that number is to 1, the higher the correlation. You can also graphically represent the correlation between two data sets by creating a "Scatter Plot" of the two datasets.

If your data does not have high correlation, that does not necessarily mean that there was something wrong in the measurement process. It just means that one sensor picked up a different reading than the other at the same time which hints to localized disturbances in the environment. Thus, a lower correlation is a hint towards hidden factors affecting the PM2.5 concentration in your environment.

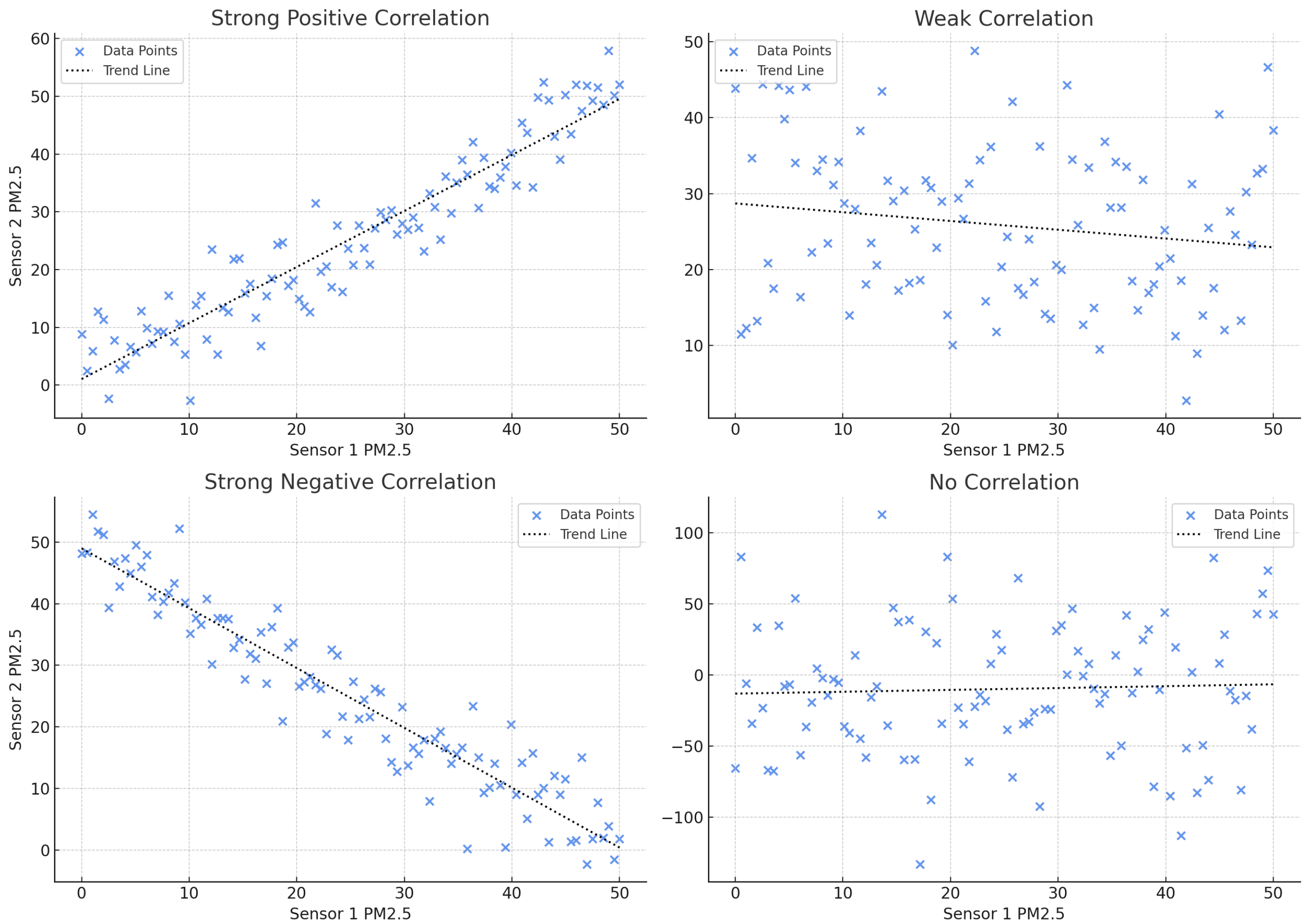


Figure 11: The scatter plots show four types of relationships: (a) strong positive correlation, (b) weak negative correlation, (c) strong negative correlation, and (d) no correlation. The black dotted line represents the best-fit trend line for each case. Strong correlations (positive or negative) indicate that the datasets change together in a predictable manner, while weak or no correlation suggests random, unrelated variations.

Note:

Correlation only defines that there is a relationship between two things, but does not always mean that one causes the other.

9. Connecting the dots

Now that we have processed our data and can actually see the changes in it with the help of a graphical plot, it's time to make sense of it. This plot really holds all the secrets within.
To fully appreciate the data, we need to look back to our surroundings and every time try to answer these questions.

- To differentiate between real changes and noise, we can compare the rolling averages of different sections of our data. If the rolling average in one section is different from another section by more than one or two times the standard deviation (σ), the change may truly be due to the environment. Otherwise, it is likely to be random noise.
- Are there any sudden changes in the concentration? If so, are these normal?
- What could contribute to a higher or lower value in the graph? If the change is real, we want to look at the "cause".

All these questions can be answered by carefully analyzing the data and linking it to possible sources of pollution or events that might have improved the air quality. By connecting patterns in the data with real-world activities around us, we can uncover the hidden stories behind the changes we observe.

Examples of Real Environmental Changes:

- Welding activities (localized increase due to particulate emission)
- Incense sticks, perfumes, or fragrances (localized chemical particulates)
- Vehicle emissions depending on fuel quality and vehicular health
- Crop burning (large-scale particulate release into the air)
- Comparison of environments near conventional brick kilns vs. zig-zag brick kilns
- Indoor kitchen activities (localized PM increase during cooking)

An example of such analysis where we identified the cause of observed value changes as well as lack of any change is as follows:

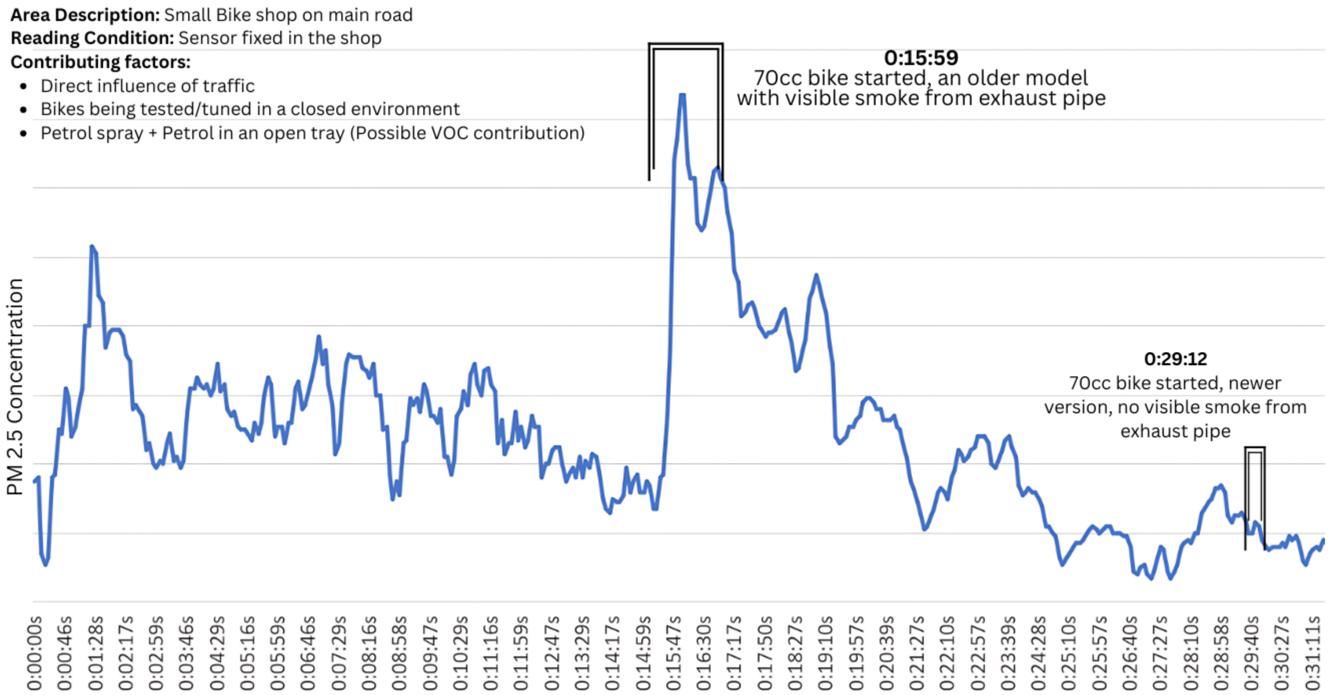


Figure 12: Spikes in the data correspond to activities such as starting older and newer motorbikes, along with background contributions from traffic and petrol handling.

10. Taking action

We have reached the end of our Phenomenon Lab. We have calculated the AQI of our environment and have analyzed our environment to observe and understand changes in the PM2.5 concentration. We've come a long way, but this is not the end. This is the time to take action on our findings. If you are conducting this experiment anywhere in Pakistan, chances are that your calculated value is more than 200.

I didn't pull a magic trick here. I know this because the impact of our destructive way of living can be seen everywhere around us. Every little selfish decision we make by being lazy or saving costs is a debt we owe to our environment. Sooner or later that debt has to be paid.
It might seem difficult, but all is not lost. Some small decisions we make individually can make a noticeable impact on the larger scale.

These actions can be:

- Stop or report any activity which emits visible smoke e.g. trash burning, small fires lit by locals etc.
- Regularly maintain our vehicles to minimize incomplete combustion in them.
- Report poorly maintained vehicles to the government [4].
- Minimize personal vehicle use by opting for public transport or walking for shorter journeys.
- Carpooling with friends to reduce vehicles on the road.

تم میں سے جو کوئی برائی دیکھے تو اسے اپنے ہاتھ سے بدل دے۔
اگر وہ ایسا نہیں کر سکتا تو اپنی زبان سے۔ اگر وہ ایسا نہیں کر
سکتا تو اپنے دل سے جو کہ ایمان کا کمزور ترین درجہ ہے۔
- الحدیث

Whoever among you sees evil, let him change it
with hid hand. If he cannot do so, then with his
tongue. If he cannot do so, then with his heart,
which is the weakest level of faith.
- Al Hadith

A. Appendix

To setup the PM Tracker, follow the step-by-step instructions in the PM Tracker User Manual.

References

[1] H. Andell. *Fine particles PM2.5 have significant impact on health.* Accessed: 2025-04-25. Nov. 7, 2022. URL: https://www.safera.com/fine-particles-pm25-are-significant-health-risk/.

[2] Dr. Sabieh Anwar. *PM Tracker - Qosain Scientific.* Accessed: 2025-04-25. Dec. 7, 2024. URL: https://qosainscientific.com/product/pm-tracker/.

[3] IQAir. *What is AQI?* Accessed: 2025-04-25. 2025. URL: https://www.iqair.com/newsroom/what-is-aqi?srsltid=AfmBOopfy9mKHs_3RiHPxfTKzj-3pLyYavLgSI-HsK1TBwUmShWohP3l.

[4] Punjab Portal. *Green Punjab | Punjab Portal.* Retrieved April 27, 2025. 2023. URL: https://punjab.gov.pk/node/6322.