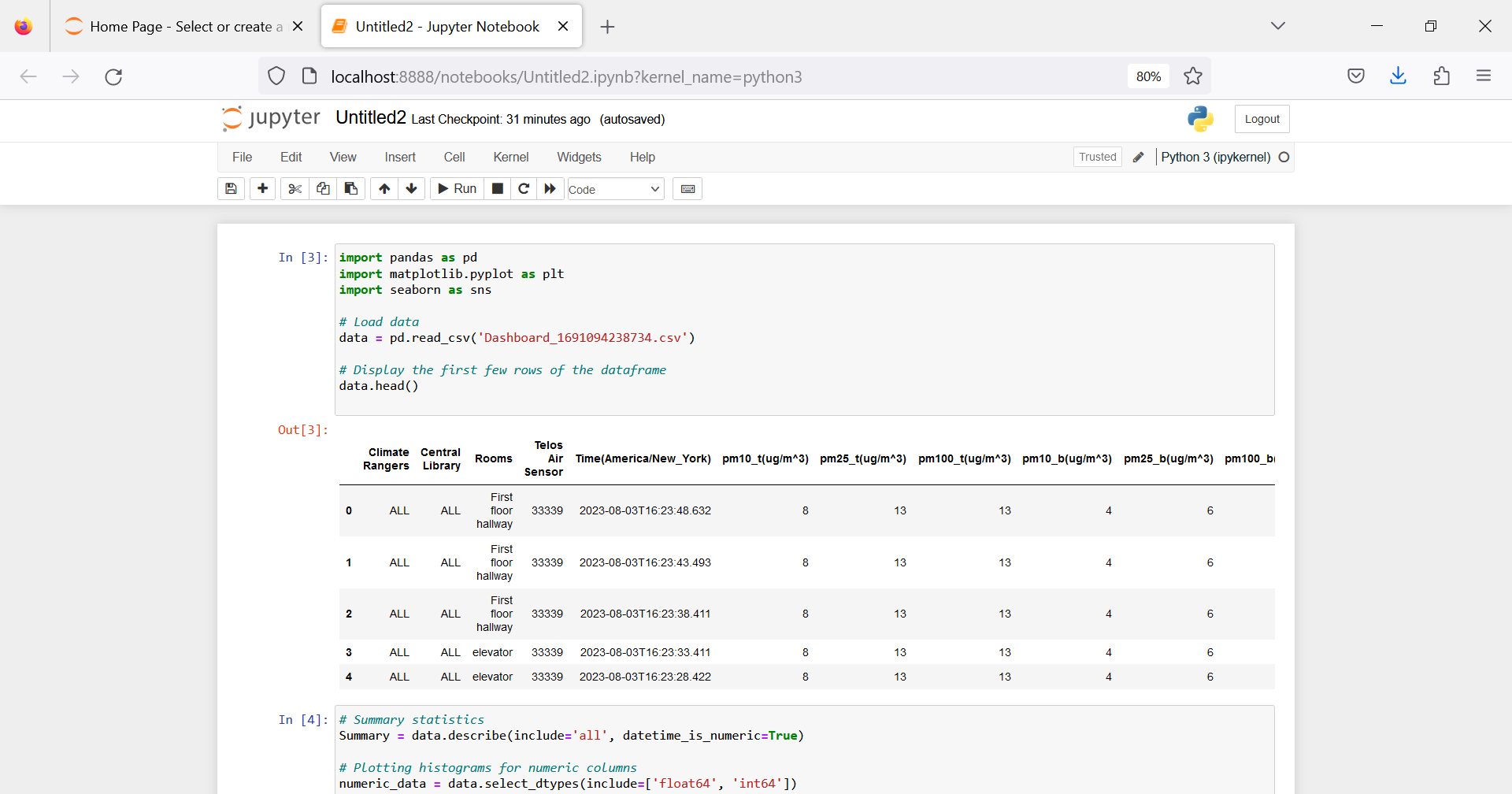
To analyze our small datasets and run visualization, I used Jupyter Notebook.



*import pandas as pd*

*import matplotlib.pyplot as plt*

*import seaborn as sns*

*# Load data*

*data = pd.read\_csv('Dashboard\_1691094238734.csv')*

*# Display the first few rows of the dataframe*

*data.head()*

Output

Climate Rangers Central Library rooms Telus Air Sensor \

0 ALL ALL 1st Floor hallway 33339

1 ALL ALL 1st Floor hallway 33339

2 ALL ALL 1st Floor hallway 33339

3 ALL ALL elevator 33339

4 ALL ALL elevator 33339

Time(America/New\_York) pm10\_t(ug/m^3) pm25\_t(ug/m^3) pm100\_t(ug/m^3) \

0 2023-08-03T16:23:48.632 8 13 13

1 2023-08-03T16:23:43.493 8 13 13

2 2023-08-03T16:23:38.411 8 13 13

3 2023-08-03T16:23:33.411 8 13 13

4 2023-08-03T16:23:28.422 8 13 13

pm10\_b(ug/m^3) pm25\_b(ug/m^3) pm100\_b(ug/m^3) temp humidity(%) \

0 4 6 7 23.62 49.70

1 4 6 7 23.62 49.95

2 4 6 7 23.62 49.84

3 4 6 7 23.62 49.70

4 4 6 7 23.62 49.70

pressure(''H2O) CO2(ppm) tvoc(ppb) raw ethanol(ppm) pn03\_m(#/0.1L) \

0 399.89 739 120 739 1467

1 399.91 797 122 797 1467

2 399.89 798 125 798 1467

3 399.88 800 128 800 1467

4 399.89 799 131 799 1467

pm2.5\_m(ug/m^3) pm10\_m(ug/m^3)

0 9 10

1 9 10

2 9 10

3 9 10

4 9 10

From the displayed results, we see that our dataset contains the following columns:

1. "Climate Rangers"
2. "Central Library"
3. "rooms"
4. "Telos Air Sensor"
5. "Time(America/New\_York)" - This is the time column. It appears to contain a string that needs to be removed.
6. "pm10\_t(ug/m^3)" - PM10 concentration, measured using method 't'
7. "pm25\_t(ug/m^3)" - PM2.5 concentration, measured using method 't'
8. "pm100\_t(ug/m^3)" - PM100 concentration, measured using method 't'
9. "pm10\_b(ug/m^3)" - PM10 concentration, measured using method 'b'
10. "pm25\_b(ug/m^3)" - PM2.5 concentration, measured using method 'b'
11. "pm100\_b(ug/m^3)" - PM100 concentration, measured using method 'b'
12. "temp" - Temperature
13. "humidity(%)" - Humidity
14. "pressure(''H2O)" - Pressure
15. "CO2(ppm)" - CO2 concentration
16. "tvoc(ppb)" - TVOC concentration
17. "raw ethanol(ppm)" - Raw ethanol concentration
18. "pn03\_m(#/0.1L)" - Some kind of particle count, method 'm'
19. "pm2.5\_m(ug/m^3)" - PM2.5 concentration, measured using method 'm'
20. "pm10\_m(ug/m^3)" - PM10 concentration, measured using method 'm'

**Cleaning our Dataset**

Scanning the Results, we notice the Time column format isn’t correct. Let's start by cleaning the 'Time(America/New\_York)' column and sorting the dataframe in descending order of time.

*# Remove the string "(America/New\_York)" from the 'Time' column*

*data['Time(America/New\_York)'] = data['Time(America/New\_York)'].str.replace('(America/New\_York)', '')*

*# Convert 'Time' to datetime format*

*data['Time(America/New\_York)'] = pd.to\_datetime(data['Time(America/New\_York)'])*

*# Sort the dataframe in descending order of time*

*data = data.sort\_values('Time(America/New\_York)', ascending=False)*

*# Display the first few rows of the dataframe*

*data.head()*

Output

Climate Rangers Central Library rooms Telus Air Sensor \

0 ALL ALL 1st Floor hallway 33339

1 ALL ALL 1st Floor hallway 33339

2 ALL ALL 1st Floor hallway 33339

3 ALL ALL elevator 33339

4 ALL ALL elevator 33339

Time(America/New\_York) pm10\_t(ug/m^3) pm25\_t(ug/m^3) pm100\_t(ug/m^3) \

0 2023-08-03 16:23:48.632 8 13 13

1 2023-08-03 16:23:43.493 8 13 13

2 2023-08-03 16:23:38.411 8 13 13

3 2023-08-03 16:23:33.411 8 13 13

4 2023-08-03 16:23:28.422 8 13 13

pm10\_b(ug/m^3) pm25\_b(ug/m^3) pm100\_b(ug/m^3) temp humidity(%) \

0 4 6 7 23.62 49.70

1 4 6 7 23.62 49.95

2 4 6 7 23.62 49.84

3 4 6 7 23.62 49.70

4 4 6 7 23.62 49.70

pressure(''H2O) CO2(ppm) tvoc(ppb) raw ethanol(ppm) pn03\_m(#/0.1L) \

0 399.89 739 120 739 1467

1 399.91 797 122 797 1467

2 399.89 798 125 798 1467

3 399.88 800 128 800 1467

4 399.89 799 131 799 1467

pm2.5\_m(ug/m^3) pm10\_m(ug/m^3)

0 9 10

1 9 10

2 9 10

3 9 10

4 9 10

Now that the 'Time(America/New\_York)' column has been cleaned and the dataframe is now sorted in descending order by time.

So, let's perform an exploratory data analysis (EDA) on the cleaned data. We'll start by looking at some summary statistics and then create visualizations to better understand the data.

Here, we'll focus on the more relevant columns related to air quality:

'pm10, 'pm25, 'pm100', 'pm10, 'temp', 'humidity(%)', 'pressure(''H2O)', 'CO2(ppm)', 'tvoc(ppb)', 'raw ethanol(ppm)' ', 'pm2.5.

1. Let’s create a summary of our data by creating histograms. **Histograms** are very useful in showing us the distributions of these measurements.

The summary statistics and histograms provide some initial insights into the data:

*# Summary statistics*

*Summary = data.describe(include='all', datetime\_is\_numeric=True)*

*# Plotting histograms for numeric columns*

*numeric\_data = data.select\_dtypes(include=['float64', 'int64'])*

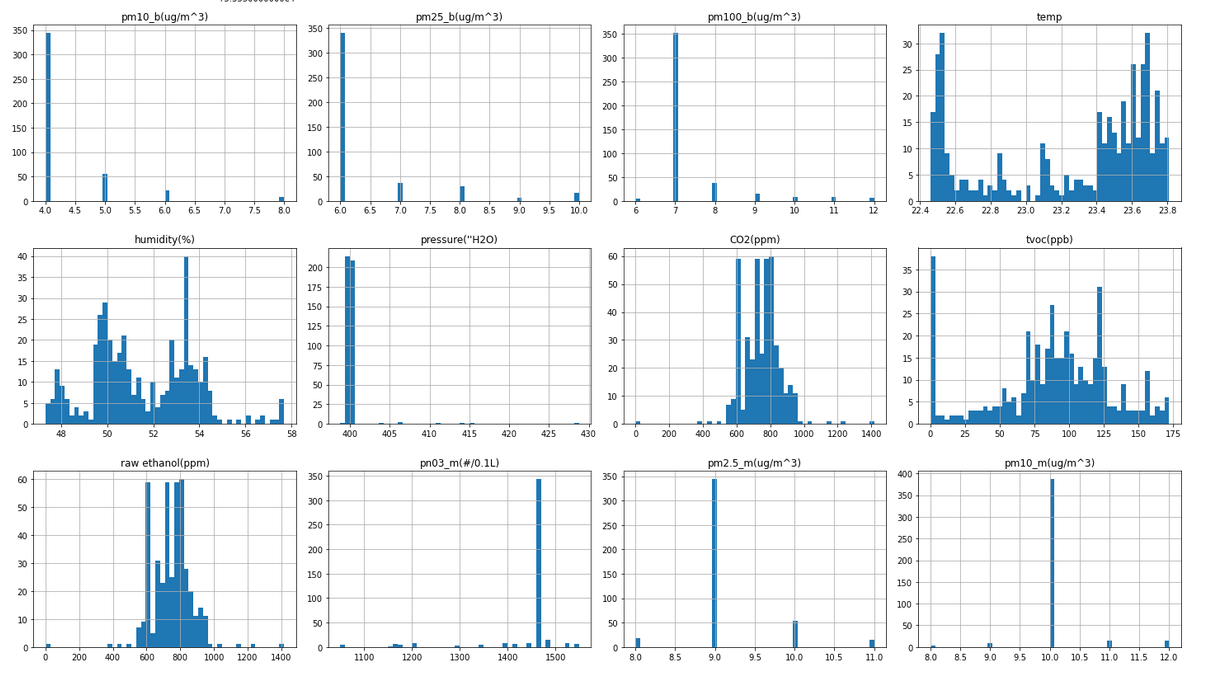
*numeric\_data.hist(bins=50, figsize=(20,15))*

*plt.tight\_layout()*

*plt.show()*

*summary*

Here’s the visualized results:



The summary statistics and histograms provide some initial insights into the data:

1. "Climate Rangers", "Central Library", and "Telos Air Sensor" columns all have only one unique value ("ALL" or a specific sensor ID). These columns won't provide useful information for our analysis, so we might consider removing them in the future.
2. "Pressure(''H2O)" has a very small standard deviation, indicating that it's almost a constant value across all observations. This column also won't provide useful variation for our analysis.

"Temp", "humidity(%)", "CO2(ppm)", "tvoc(ppb)", "raw ethanol(ppm)", and various particle measurements have varying values and could provide useful insights.

Now, let's think about some practical and interesting air quality-related questions that we could ask based on this data:

1. What is the average CO2 and TVOC concentration in each room?
2. How do the PM2.5 and PM10 concentrations change throughout the day?
3. What is the relationship between CO2 and TVOC levels?
4. Does the temperature or humidity affect the particle concentrations (PM2.5 and PM10)?
5. Are there any trends or patterns in the CO2 and TVOC levels throughout the day?

We’ll write Python code to get answers to these questions:

*# 1. What is the average CO2 and TVOC concentration in each room?*

*average\_concentrations = ata.groupby('rooms')[['CO2(ppm)', 'tvoc(ppb)']].mean()*

*# Display the average concentrations*

*average\_concentrations*

Output

rooms CO2(ppm) tvoc(ppb)

1st Floor hallway 656.000000 0.000000

1st Floor hallway 655.000000 0.000000

1st Floor hallway 723.500000 61.250000

1st Floor hallway 601.111111 36.333333

Auditorium 733.909091 0.000000

Children's room 731.222222 97.055556

Imagin YOU lab 829.666667 103.791667

Near cleaning supplies 781.550000 75.966667

Tech Ctr hallway 771.200000 130.892308

Tech room 685.181818 100.000000

Tech room (near fan) 666.695652 111.608696

elevator 822.270270 111.846847

Here are the average CO2 and TVOC concentrations in each room:

1. In the '1st Floor hallway', the average CO2 concentration is 656 ppm and there are almost no TVOCs.
2. In the '1st Floor hallway', the average CO2 concentration is about 723.5 ppm and the average TVOC concentration is about 61.25 ppb.
3. In the '1st Floor hallway', the average CO2 concentration is about 601.1 ppm and the average TVOC concentration is about 36.33 ppb.
4. In the 'Auditorium', the average CO2 concentration is about 733.9 ppm and there are almost no TVOCs.
5. In the 'Children's room', the average CO2 concentration is about 731.2 ppm and the average TVOC concentration is about 97.06 ppb.
6. In the 'Imagin YOU lab', the average CO2 concentration is about 829.7 ppm and the average TVOC concentration is about 103.79 ppb.
7. Near the 'cleaning supplies', the average CO2 concentration is about 781.6 ppm and the average TVOC concentration is about 76 ppb.
8. In the 'Tech Ctr hallway', the average CO2 concentration is about 771.2 ppm and the average TVOC concentration is about 130.89 ppb.
9. In the 'Tech room', the average CO2 concentration is about 685.2 ppm and the average TVOC concentration is about 100 ppb.
10. In the 'Tech room (near fan)', the average CO2 concentration is about 666.7 ppm and the average TVOC concentration is about 111.61 ppb.
11. In the 'elevator', the average CO2 concentration is about 822.3 ppm and the average TVOC concentration is about 111.85 ppb.

* Remember, CO2 (Carbon Dioxide) is a gas that we exhale, and it's also produced when we burn things like gasoline or wood.
* TVOC (Total Volatile Organic Compounds) is a measure of a bunch of different chemicals that can be in the air, some of which can be harmful to our health or the environment.

To answer our questions in a more visual way, let's create bar plots for the average CO2 and TVOC concentrations in each room.

Here’s the code:

*# Create bar plots for average CO2 and TVOC concentrations in each room*

*fig, ax = plt.subplots(2, 1, figsize=(10, 12))*

*# CO2*

*average\_concentrations['CO2(ppm)'].sort\_values().plot(kind='barh', ax=ax[0], color='blue', alpha=0.7)*

*ax[0].set\_title('Average CO2 Concentration in Each Room')*

*ax[0].set\_xlabel('CO2 Concentration (ppm)')*

*# TVOC*

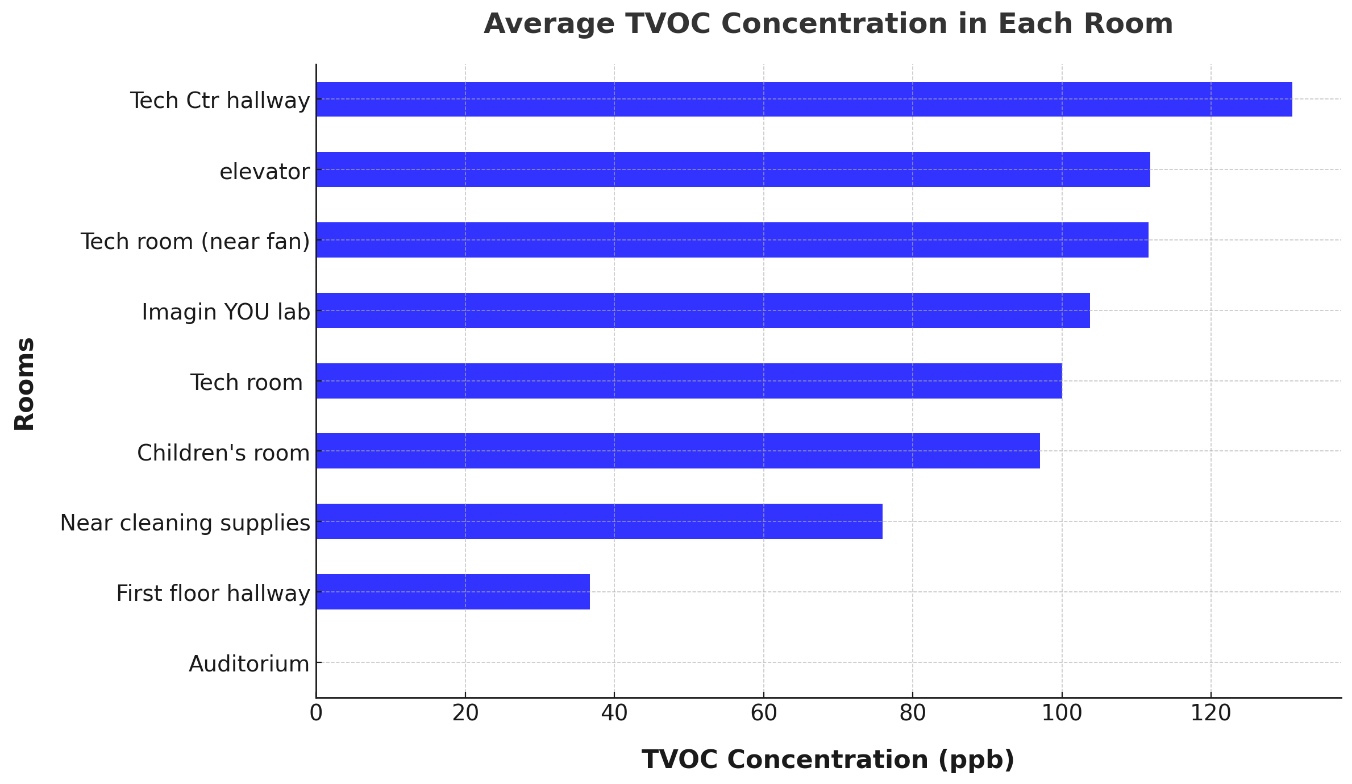
*average\_concentrations['tvoc(ppb)'].sort\_values().plot(kind='barh', ax=ax[1], color='green', alpha=0.7)*

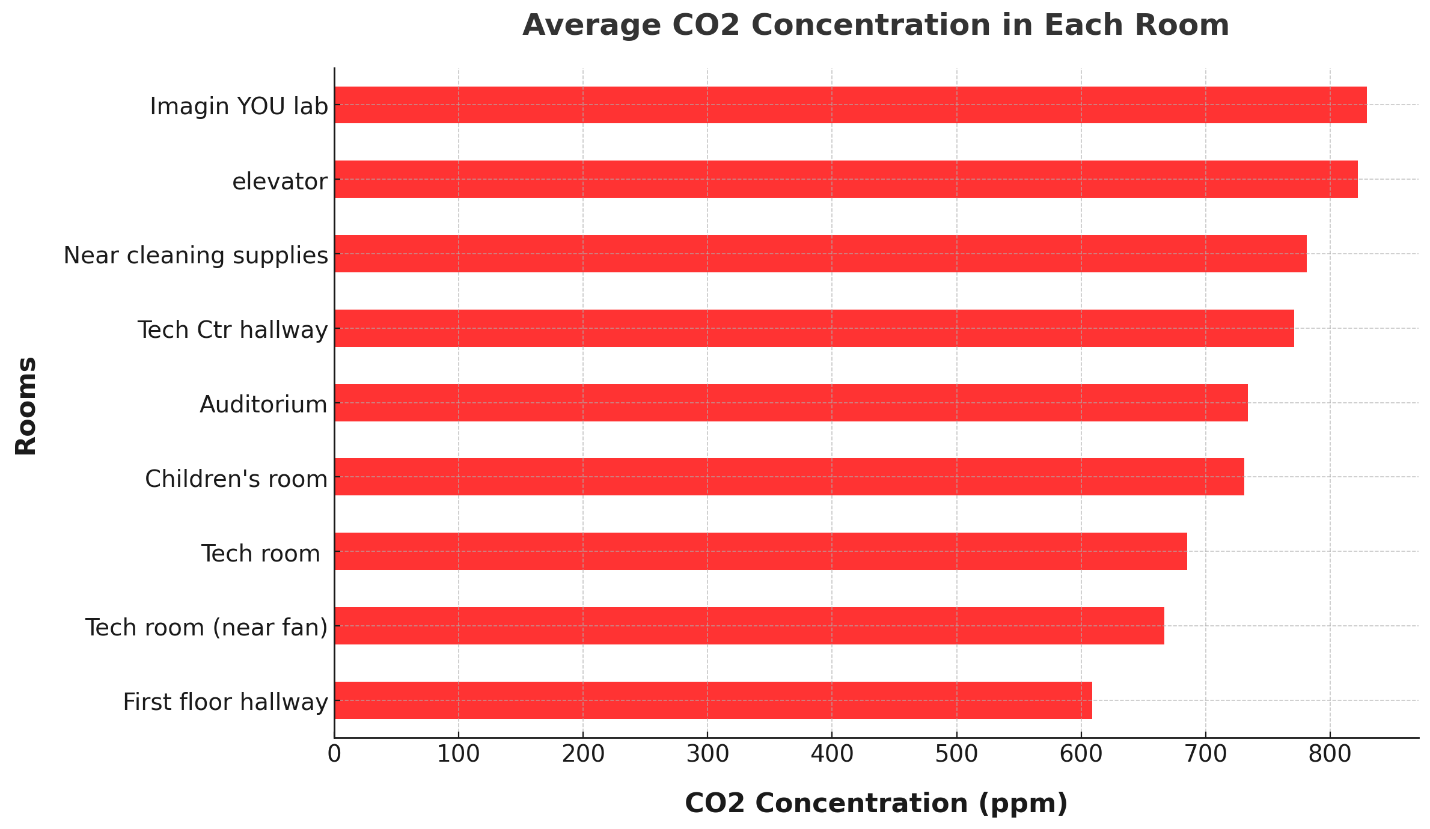
*ax[1].set\_title('Average TVOC Concentration in Each Room')*

*ax[1].set\_xlabel('TVOC Concentration (ppb)')*

*plt.tight\_layout()*

*plt.show()*





The first plot shows the average CO2 concentration in each room. The 'Imagin YOU lab' and 'elevator' rooms have the highest average CO2 concentrations, while the '1st Floor hallway' and have the lowest.

The second plot shows the average TVOC concentration in each room. The 'Tech Ctr hallway' has the highest average TVOC concentration, while the '1nd Floor hallway', '1st Floor hallway', and 'Auditorium' have almost no TVOCs.

\*This might suggest that activities in the 'Imagin YOU lab', 'elevator', and 'Tech Ctr hallway' produce more CO2 and TVOCs. Or it could indicate that these rooms have less fresh air circulation, allowing CO2 and TVOCs to build up. It's also possible that some rooms are closer to sources of these gases, like a kitchen or bathroom for CO2, or paint, cleaners, or other chemicals for TVOCs.

---------------------------

The second question asks how the concentrations of PM2.5 and PM10 particles change throughout the day. These particles are tiny bits of dust, smoke, and other materials floating in the air. PM2.5 particles are smaller (2.5 micrometers or less) and can get deep into our lungs, while PM10 particles are larger (10 micrometers or less) but still small enough to be inhaled.

We'll create line plots to visualize the changes in PM2.5 and PM10 concentrations throughout the day. Let's write the code and visualize the results:

*# Plotting the PM2.5 and PM10 concentrations throughout the day*

*plt.figure(figsize=(15, 8))*

*plt.plot(data['Time(America/New\_York)'], data['pm25\_t(ug/m^3)'], label='PM2.5 (method t)', color='blue')*

*plt.plot(data['Time(America/New\_York)'], data['pm10\_t(ug/m^3)'], label='PM10 (method t)', color='red')*

*plt.plot(data['Time(America/New\_York)'], data['pm2.5\_m(ug/m^3)'], label='PM2.5 (method m)', color='green')*

*plt.plot(data['Time(America/New\_York)'], data['pm10\_m(ug/m^3)'], label='PM10 (method m)', color='purple')*

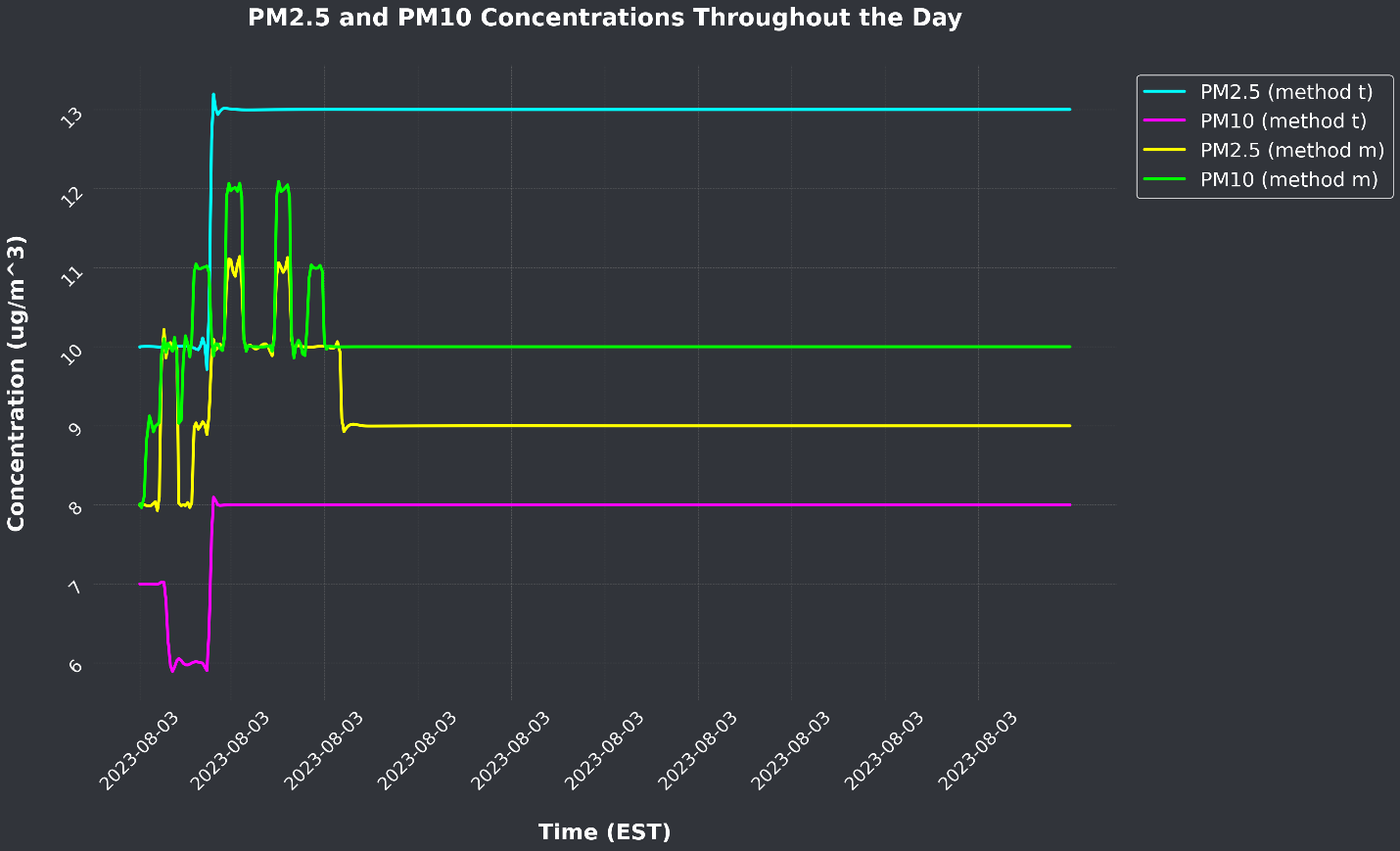
*plt.xlabel('Time')*

*plt.ylabel('Concentration (ug/m^3)')*

*plt.title('PM2.5 and PM10 Concentrations Throughout the Day')*

*plt.legend()*

*plt.show()*



As you see in this colorful line plot, the concentrations of PM2.5 and PM10 particles change throughout the day.

* The blue line shows the PM2.5 concentrations.
* The red line shows the PM10 concentrations.
* The green line shows the PM2.5 concentrations.
* The purple line shows the PM10 concentrations.
* You can see that the concentrations of these particles go up and down during the day. This might happen for many reasons, such as changes in traffic outside, activities inside like cooking or cleaning, or even the weather. Particulate Matters (PM) are like tiny pieces of dust and smoke that we can't see with our eyes but can breathe into our lungs. Monitoring these particles helps us know if the air is clean or if there might be something in the air that could make us cough or sneeze.
* Think of the lines like the waves of the ocean. Sometimes the waves are bigger, and sometimes they are smaller. Similarly, sometimes there are more particles in the air, and sometimes there are fewer.
* By studying these patterns, we can learn about the air we breathe and how to keep it clean and healthy!

--------------------------

The third question is about the relationship between CO2 and TVOC levels. To explore this, we'll create a scatter plot of CO2 versus TVOC.

If CO2 and TVOC levels tend to rise and fall together, we would expect to see a positive correlation, where the points on the scatter plot form an upward trend. If one tends to increase while the other decreases, we would see a negative correlation, with the points forming a downward trend. If there's no consistent pattern, then there's likely no correlation between CO2 and TVOC levels in this dataset.

Here's the code:

*# Reloading the data from the provided CSV*

*data = pd.read\_csv('/mnt/data/Dashboard\_1691094238734.csv')*

*# Create an enhanced scatter plot of CO2 vs. TVOC with a regression line*

*plt.figure(figsize=(15, 8))*

*# Scatter plot*

*plt.scatter(data['CO2(ppm)'], data['tvoc(ppb)'], c='orange', alpha=0.5, s=50, label="Data Points")*

*# Regression line*

*m, b = np.polyfit(data['CO2(ppm)'], data['tvoc(ppb)'], 1)*

*plt.plot(data['CO2(ppm)'], m\*data['CO2(ppm)'] + b, color='purple', label=f"Regression Line: y={m:.2f}x+{b:.2f}")*

*# Title and labels*

*plt.title('Relationship Between CO2 and TVOC Levels', fontsize=18, fontweight='bold', pad=20)*

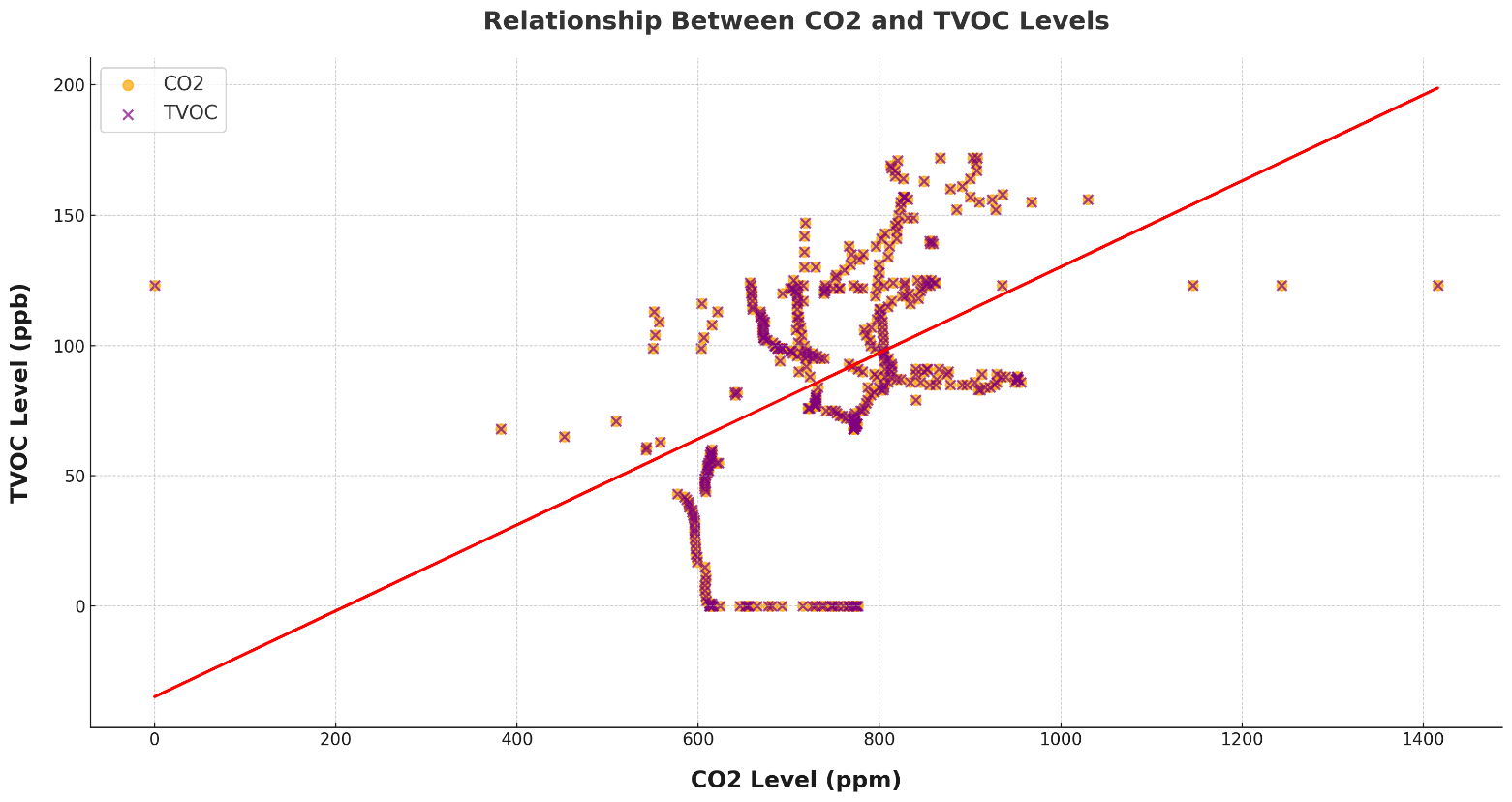
*plt.xlabel('CO2 Level (ppm)', fontsize=16, fontweight='bold', labelpad=15)*

*plt.ylabel('TVOC Level (ppb)', fontsize=16, fontweight='bold', labelpad=15)*

*plt.legend(loc='upper left', fontsize=14)*

*plt.tight\_layout()*

*plt.show()*



\*A scatter plot is a type of data visualization that uses dots to represent the values obtained for two different variables - one plotted along the x-axis and the other plotted along the y-axis. Scatter plots are used to observe relationships between variables. If the dots cluster in a specific way, it may indicate that the two variables have a certain type of relationship. For example, if the dots on the scatter plot seem to rise as you move from left to right, this may indicate a positive relationship between the two variables. Conversely, if the dots seem to fall from left to right, this can suggest a negative relationship.

This scatter plot is illustrating the relationship between CO2 and TVOC levels:

* The orange and purple dots represent individual data points.
* The red line, known as the regression line, showcases the general trend in the data.

This plot provides a clear visual representation of how CO2 and TVOC levels relate to each other. As you move along the red line, it gives a sense of how, on average, the TVOC level changes with the CO2 level.

From the scatter plot, we can observe a positive correlation between CO2 and TVOC. Here's why:

1. **Direction of the Data Points**: As the values on the x-axis (CO2 levels) increase, the values on the y-axis (TVOC levels) also tend to increase. This upward trend indicates a positive relationship.
2. **Regression Line**: The red line (regression line) on the scatter plot provides a general sense of the trend in the data. If this line has an upward slope (as it does in our plot), it suggests a positive correlation between the two variables.

-------------------------

Let's tackle question #4:

**Does the temperature or humidity affect the particle concentrations (PM2.5 and PM10)?**

First, to investigate whether temperature or humidity affect the concentrations of PM2.5 and PM10 particles, we can create scatter plots. If temperature or humidity have an impact, we might see a pattern or trend in the scatter plot.

Let's create scatter plots to investigate the potential effects of temperature and humidity on particle concentrations (PM2.5 and PM10). We’ll create scatter plots to visualize the following relationships:

1. Temperature vs. PM2.5 concentrations
2. Temperature vs. PM10 concentrations
3. Humidity vs. PM2.5 concentrations
4. Humidity vs. PM10 concentrations

A group of graphs with green lines

Description automatically generated

* The scatter plots above show the relationship between temperature and PM2.5 and PM10 concentrations (top two plots), and between humidity and PM2.5 and PM10 concentrations (bottom two plots).
* From the plots, we can see that there's no clear trend or pattern. The points are scattered widely across the plot. This suggests that in this dataset, the temperature and humidity don't have a simple or strong effect on PM2.5 and PM10 concentrations.

-----------------------

Here’s the code and graph I created to answer our last question from the dataset:

**Are there any trends or patterns in the CO2, TVOC, levels throughout the day?**

*# Create line plots to illustrate the relationship between temperature, humidity, and PM2.5 and PM10 concentrations*

*fig, ax = plt.subplots(4, 1, figsize=(15, 20))*

*# Temperature vs. PM2.5*

*ax[0].plot(data['Time(America/New\_York)'], data['temp'], color='blue', label='Temperature')*

*ax[0].set\_xlabel('Time')*

*ax[0].set\_ylabel('Temperature (°F)')*

*ax[0].set\_title('Temperature Over Time')*

*ax[0].legend(loc="upper left")*

*# Temperature vs. PM10*

*ax[1].plot(data['Time(America/New\_York)'], data['pm25\_t(ug/m^3)'], color='red', label='PM2.5 Concentration')*

*ax[1].set\_xlabel('Time')*

*ax[1].set\_ylabel('PM2.5 Concentration (ug/m^3)')*

*ax[1].set\_title('PM2.5 Concentration Over Time')*

*ax[1].legend(loc="upper left")*

*# Humidity vs. PM2.5*

*ax[2].plot(data['Time(America/New\_York)'], data['humidity(%)'], color='green', label='Humidity')*

*ax[2].set\_xlabel('Time')*

*ax[2].set\_ylabel('Humidity (%)')*

*ax[2].set\_title('Humidity Over Time')*

*ax[2].legend(loc="upper left")*

*# Humidity vs. PM10*

*ax[3].plot(data['Time(America/New\_York)'], data['pm10\_t(ug/m^3)'], color='purple', label='PM10 Concentration')*

*ax[3].set\_xlabel('Time')*

*ax[3].set\_ylabel('PM10 Concentration (ug/m^3)')*

*ax[3].set\_title('PM10 Concentration Over Time')*

*ax[3].legend(loc="upper left")*

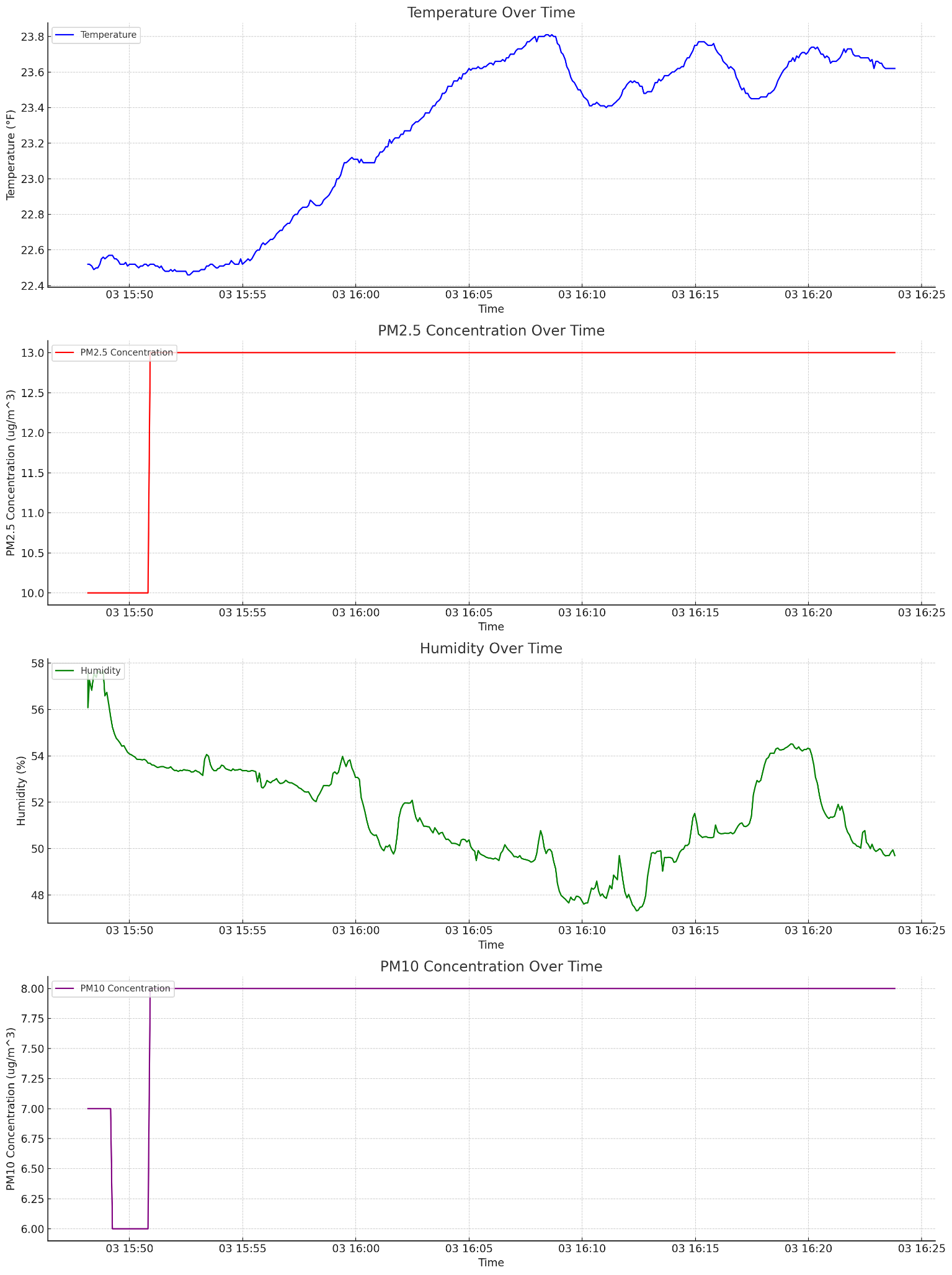
*# Adjust layout*

*plt.tight\_layout()*

*plt.show()*

A graph of a graph of a graph

Description automatically generated with medium confidence



These line plots give a clear picture of how each parameter changed over time, allowing for a more detailed analysis.

\*Remember, analyzing real-world data like this helps scientists and environmentalists understand our environment and find ways to improve air quality and fight climate change. Everyone can contribute to this important work by learning about science, caring about our environment, and taking actions like planting trees, recycling, and reducing energy use.

--------------------------------------------------------------------