# **Air Quality Prediction Analysis of Kathmandu City: Impact of Environmental Factors in the Suspended Particulate Matters**

*Author: Ganesh Sapkota*

*Date: 01/19/2021*

#### **Problem: Air Pollution in Kathmandu**

Kathmandu, the capital city of Nepal is one of the fastest-growing cities in south Asia which is situated within a valley and many mountains ranges around. Due to rapid growth in construction-related activities, vehicles and industries, air pollution has been a quite prominent the problem in this city. In 2019, Kathmandu was ranked as the seventh most polluted capital city of the world in terms of Air quality index (AQI), according to research by IQ Air Visual, a Swiss-based group that gathers air-quality data globally. As of 4th January 2021, Kathmandu recorded highest AQI index of 450. The key factors of air pollution in Kathmandu might be emissions from old and outdated vehicles relying on diesel fuels, open burn fires of organic garbage, dust, and finely ground particles from a construction sites compounded by its geographical location, lacking the elevation and wind to allow these pollutants to disperse properly, instead accumulating and rising to dangerous levels.

#### **Major Pollutants**

A major source of pollutants came from burning and combustion, in the form of fine particulate matter (PM) of black carbon. Particulate matter PMx is categorized based on its diameter x. The most considerable PMx are PM2.5 and PM10. PM2.5 indicates fine particles with a diameter less than or equal to 2.5 micrometers while PM10 refers to coarse dust particles having a diameter of 2.5 to 10 micrometers. In 2019 Kathmandu recorded PM2.5 reading of 48 μg/m³ as a yearly average, placing it into the ‘unhealthy for sensitive groups’ bracket, which requires a PM2.5 reading of anywhere between 35.5 to 55.4 μg/m³.Other pollutants arising from vehicles would include carbon monoxide (CO), nitrogen dioxide (NO2), ozone (O3), and sulfur dioxide (SO2).

#### **Effect of Environmental Factors in PM10 Level**

The weather factors play an important role in the transport, diffusion and distributions of air pollution. The entering of pollutants from the ground surface, their residence in the atmosphere, and the formation of secondary pollutants only depends not on the only rate of emission of the reactants into the air from the source, but might also on wind speed, the direction of the wind, weather temperature, relative humidity.

Thus, it is often important to understand their behavior leading to an observed concentration of pollutants at a given point.

The following piece of work is done with an intention to perform statistical analysis on Suspended Particulate Matters and different environmental factors in Kathmandu city and to develop insight by observing their separate relationships with each other. The exercise has been organized as A) Data Pre-Processing and B) Statistical Analysis.

1. **Data Pre-Processing**

**Original Dataset Sample:**

|  | **Unnamed: 0** | **T\_x** | **datetime** | **station\_name** | **PM1\_I** | **PM10\_I** | **PM2.5\_I** | **RH\_I** | **WS\_I** | **WD\_I** | **TSP\_I** | **long** | **latt** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 0 | 15.9 | 2020-10-26 00:01:00+00:00 | Nepalgunj | 23.100000 | 23.400000 | 23.400000 | 92.900002 | 0.7 | 302.799988 | 23.400000 | 81.6222 | 28.05275 |
| **1** | 1 | 15.9 | 2020-10-26 00:02:00+00:00 | Nepalgunj | 22.299999 | 22.299999 | 22.299999 | 92.900002 | 0.5 | 278.100006 | 22.299999 | 81.6222 | 28.05275 |
| **2** | 2 | 15.9 | 2020-10-26 00:03:00+00:00 | Nepalgunj | 21.900000 | 21.900000 | 21.900000 | 93.099998 | 0.5 | 290.100006 | 21.900000 | 81.6222 | 28.05275 |
| **3** | 3 | 15.8 | 2020-10-26 00:04:00+00:00 | Nepalgunj | 21.600000 | 22.000000 | 21.900000 | 93.199997 | 0.3 | 299.000000 | 22.000000 | 81.6222 | 28.05275 |
| **4** | 4 | 15.8 | 2020-10-26 00:05:00+00:00 | Nepalgunj | 22.200001 | 22.200001 | 22.200001 | 93.300003 | 0.4 | 323.200012 | 22.200001 | 81.6222 | 28.05275 |

The dataset has 13 columns and 158502 data points which indicate the data associated with different parameters collected by respective sensors at different stations every second. The columns appearing in the datasets that will behave as different features/variables in statistical analysis below are briefly explained here:

**Unnamed:** Index of datapoints

**datetime:** Timestamp of data collection by sensors

**Station\_name:** Name of the place/city where sensors are placed

**long:** Longitude

**latt:** Latitude

**Pollutant Variables:**

**PM1**: Fine particles with diameter less than or equal to 1 micrometer measured as µg/m3

**PM2.5**: Fine particles with a diameter less than or equal to 2.5 micrometers measured as µg/m3

**PM10**: Coarse dust particles having a diameter of 2.5 to 10 micrometers measured as µg/m3

**TSP**: Total Suspended Particle, indicates mass concentration of particulate matter (PM) in community air measured as µg/m3

**Environmental variables:**

**T\_x**: Weather Temperature measured in degree Celsius

**RH**: Relative Humidity

**WS**: Wind Speed measured as Knot (nautical mile per hour = 0.51 m sec-1 = 1.15 mph)

**WD**: Wind Direction measured in Degrees

**Final Dataset sample (Pre-Processed Dataset)**

|  | **Temp** | **PM10\_I** | **PM2.5\_I** | **Rel\_Humid** | **Wind\_speed** | **Wind\_dir** | **tot\_sus\_particle** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **datetime** |  |  |  |  |  |  |  |
| **2020-10-26 00:00:00+00:00** | 16.500000 | 11.066667 | 8.857576 | 90.896971 | 0.400000 | 182.381818 | 13.684848 |
| **2020-10-26 01:00:00+00:00** | 17.077419 | 15.738710 | 11.758064 | 89.020968 | 0.670968 | 176.798387 | 19.333871 |
| **2020-10-26 02:00:00+00:00** | 18.963158 | 57.133332 | 21.519298 | 82.547368 | 0.792982 | 175.903509 | 113.636843 |
| **2020-10-26 03:00:00+00:00** | 21.137288 | 54.981356 | 13.228814 | 74.181356 | 0.906780 | 197.257627 | 108.900000 |
| **2020-10-26 04:00:00+00:00** | 22.665116 | 47.718604 | 12.355814 | 66.858140 | 0.976744 | 195.265118 | 101.720929 |

# **Statistical Analysis**

### **Basic statistics of variables under consideration**

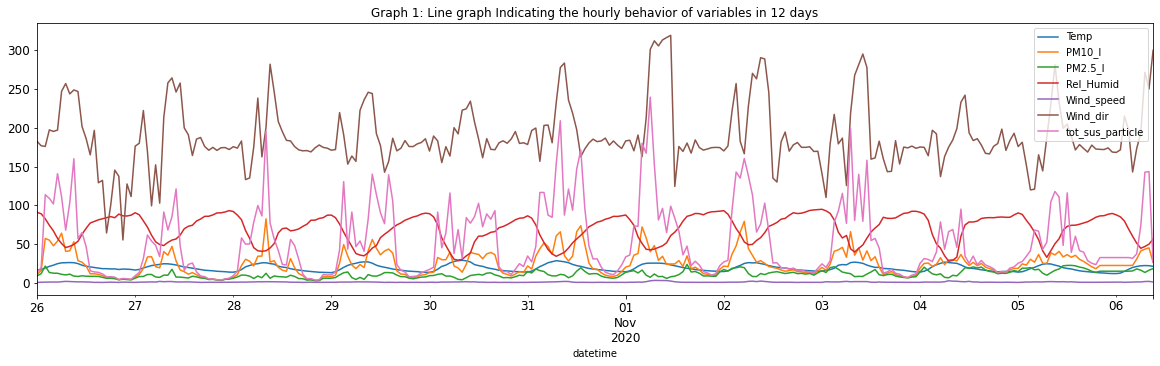
|  | **Temp** | **PM10\_I** | **PM2.5\_I** | **Rel\_Humid** | **Wind\_speed** | **Wind\_dir** | **tot\_sus\_particle** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | 274.000000 | 274.000000 | 274.000000 | 274.000000 | 274.000000 | 274.000000 | 274.000000 |
| **mean** | 19.328189 | 25.060348 | 10.986991 | 69.692333 | 0.903694 | 188.962409 | 52.980355 |
| **std** | 4.300853 | 15.272235 | 4.543881 | 17.881896 | 0.472779 | 41.694798 | 45.956292 |
| **min** | 11.868750 | 2.932000 | 2.576000 | 27.772727 | 0.400000 | 55.228813 | 3.374545 |
| **25%** | 15.425522 | 13.194857 | 7.613362 | 55.634860 | 0.575102 | 170.810013 | 16.549708 |
| **50%** | 18.487807 | 22.299999 | 10.184405 | 74.431555 | 0.739167 | 179.411744 | 34.365216 |
| **75%** | 22.886538 | 34.430877 | 13.902753 | 84.893212 | 1.062586 | 198.748805 | 80.125581 |
| **max** | 28.967308 | 82.462963 | 23.746667 | 94.791667 | 3.018519 | 319.227273 | 239.453966 |

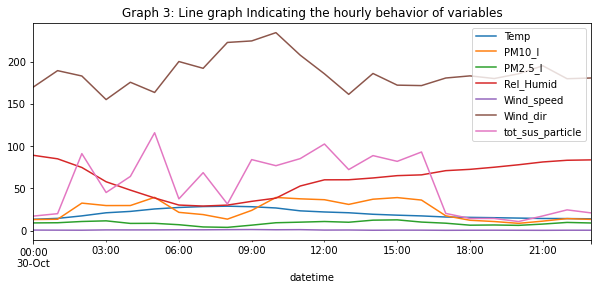
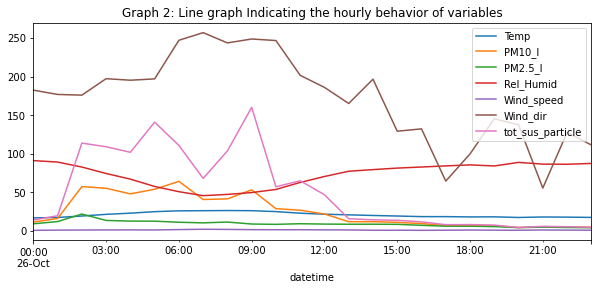
##### Summary of Statistics:

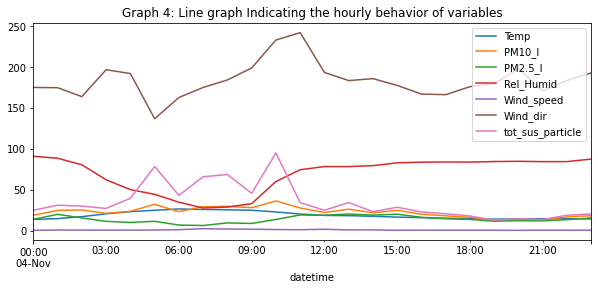
According to the above table, PM2.5 ranged between 2.57 to 23.31 with a mean of 10.98 while PM10 ranged between 2.93 and 82.46 with the mean of 25.06 in 12 days starting from the 26th of October 2020 to the 6th of November 2020. Average values of PM2.5 and PM10 in 12 days indicate that both the levels are within the Satisfactory scale ie(0-50).

Also, the TSP (Total Suspended Particle) ranges from 3.37 to 239.45 with the mean of 52.98 within this period. it indicates that the mean value falls in the moderate scale. TSP show a large value of the range within this period with a significant standard deviation of 45.95 which is the highest compared to other pollutants

### **2. Graph Analysis**



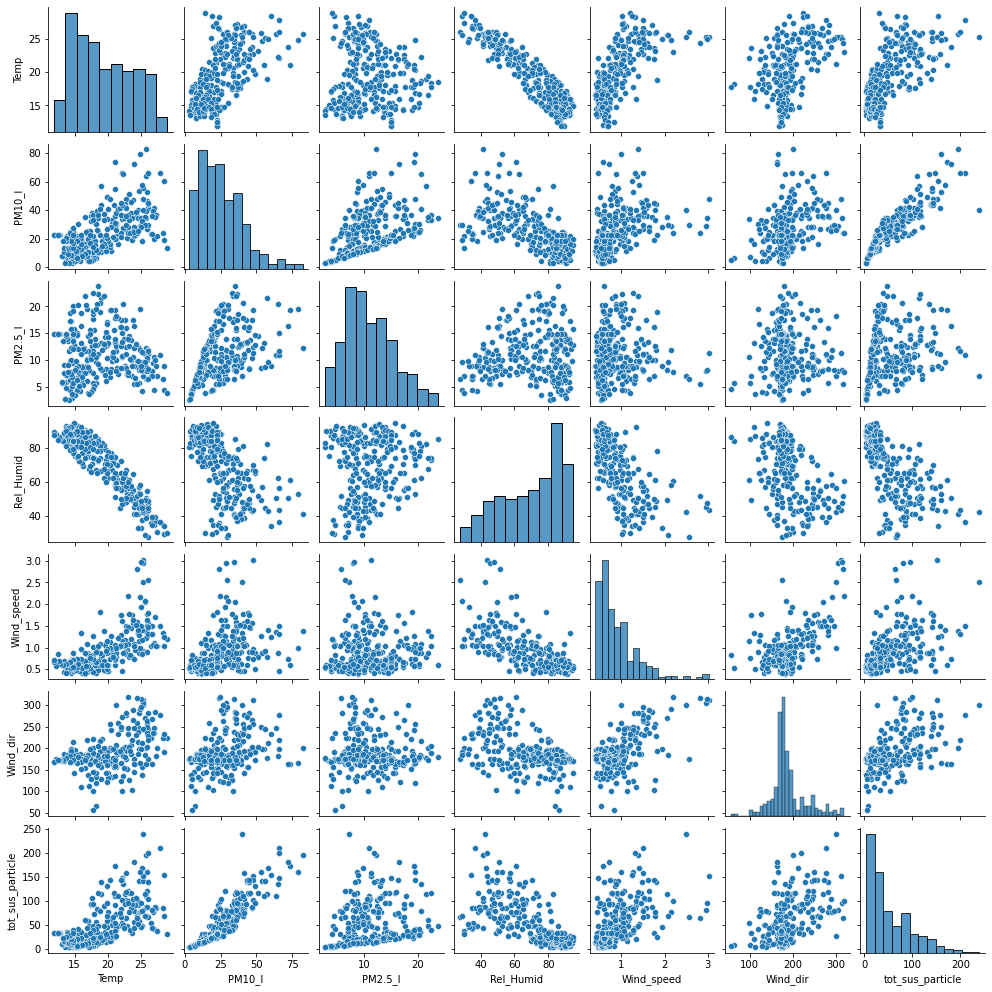




Above line graphs help to understand and analyze the historical readings(values) of pollutant variables and weather variables against date time. Graph 1 shows an hourly average reading of the total time frame of 12 days on which readings were done. It shows an almost similar pattern of rising and fall in each day throughout the period.

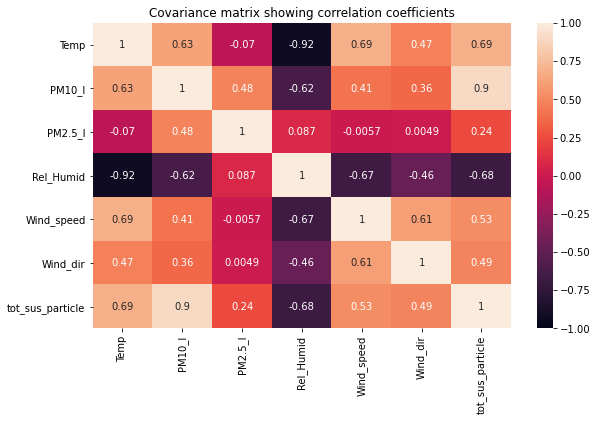
Graph 2, Graph 3, and Graph 4 shows the behavior of variables in three random days - ie. 26th Oct, 30th Oct, and 4th Nov. The graph of three individual days, helps to understand a general pattern of rising and fall of the variables' reading in 24 hours through out a day.

1. **Pair Plot Analysis:**



The above pair plot displays the association between pollutants like PM2.5, PM10, and TSP with different environmental factors like temperature, relative humidity, wind speed, and wind direction. It shows some significant association of PM10 with TSP, relative humidity, and temperature. PM2.5 also seems to have some association to TSP but its association with other factors looks weaker. we can also figure out some association between PM2.5 and PM10 as well. TSP itself shows a different level of association with wind speed, wind direction and temperature.

### **Calculation of covariance matrix**



### **Observations and Insights**

From the covariance matrix plot above based on Pearson's correlation coefficient, we can see different levels of correlations between environmental variables and weather variables. PM10 shows a weak positive correlation with wind direction and wind speed but a moderate positive correlation with temperature while it shows a moderate negative correlation with relative humidity. it means PM10 level might rise with rising in temperature, wind speed, and wind turbulence level and vice versa while its concentration might fall with rising in relative humidity level and vice versa.

The PM2.5 doesn't show any significant correlation with any of the weather variables. On the other hand, the TSP level shows a moderate correlation with environmental variables which is quite interesting. It shows a positive moderate correlation with wind direction, wind speed, and temperature while it is observed that relative humidity inversely correlates with TSP with the moderate value of correlation coefficient.

The correlation of TSP with PM10 is quite strong positive but that with PM2.5 is weak positive. It can also be observed that PM10 moderately correlates with PM2.5.

From the above analysis, it is observed that TSP - A Particulate matter which indicates the sum of all solid and liquid particles suspended in the air that contributes to air pollution, significantly correlates with weather variables - Temperature, Wind speed, Wind Direction, Relative Humidity and with Pollutant variables - PM10 and PM2.5. So, the level of TSP (Output Variable) can be Predicted using the model that Implement algorithms like Linear Regression utilizing other variables as an input. Prediction of PM10 can also be modeled in the same fashion.

But the case of PM2.5 is different. Although it shows some correlation with PM10 and TSP, its functional relationship with weather variables is not trivial. So, it might require some complex approaches to predict the influence of weather variables on PM2.5 levels in the air.

By analyzing the above association of pollutant variables with environmental/meteorological variables, we can develop an insight that observations of environmental/meteorological variables might play important role in predicting/forecasting the level of pollutants in Kathmandu city. Similarly, the measurement of TSP level might also help predict the level of other pollutants like PM10 and PM2.5.