INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume3, Issue4)
Available online at www.ijariit.com

Analysis of Air Pollution

Rajeshwari K. Rai
BITS Pilani, India
krrai77@gmail.com

Abstract: This research paper is an attempt towards analyzing real time air pollution data collected by PAQS sensor devices from some key locations in Bangalore. Air pollution in most of the metropolitan cities in India is turning out to be a major threat to our environment and hazardous to our health. Many infections and diseases related to lungs and throat are caused by the polluted air we breathe. There is a growing need to conduct regular measurements of air quality data and analyze it.

Keywords: Air Pollution, IOT Sensors, Environment, Cities

I. INTRODUCTION

Air pollution in India is emerging as a major factor that seems to contribute to many of the health hazards that people are facing such as asthma, lung and chest infections, and allergies. With the increasing number of vehicles used in metropolitan cities the quality of air that we breathe every second is turning out to be very harmful to our health. This has led to the growing need to conduct periodic measurements of air quality data especially the calculation of air pollutant concentrations along with other data captured at various places and time zones and analysis of it can help to understand the pointers such as the major contributors of air pollution in a particular location, what kind of air pollution carries more hazardous air pollutant, during what time or season the air pollutants are more active, etc. These important resultant data will, in turn, help people to know the areas that are more prone to particular air pollutants and the source of air pollution that is causing this. This data will also help environmentalist to work towards their goal of maintaining ecological balance.

II. THEORETICAL BACKGROUND

Through air pollution, harmful substances are released into the atmosphere. This in-turn causes diseases and sometimes even death in humans. It also harms animals and food crops. Also, air pollution has an adverse effect on our environment. In urban areas, the main cause of air pollution is traffic, industries, and factories. Air pollution through these mediums is caused by the substances (pollutants) released into the air that results in severe damage to the ecosystem [1].

Pollutants are classified as primary or secondary. Primary pollutants are usually produced from a process, examples include carbon monoxide gas from motor vehicle exhaust or the sulfur dioxide released from factories. Secondary pollutants are not emitted directly, they form in the air when primary pollutants react or interact. O3 is a prominent example of a secondary pollutant. Some pollutants may be both primary and secondary. The level of air pollutants in the ambient air is affected by factors such as atmospheric wind speed, wind direction, relative humidity, and temperature. RH depends on temperature inversely. It requires less water vapor to attain high relative humidity at low temperatures and more water vapor is required to attain high relative humidity in the warm or hot air[1].

PAQS provide context aware IOT solutions from secured real time monitoring wearable gadgets to individuals, Smart Homes & Offices, Smart cities, and Community. PAQS solutions enable various customers to monitor the inhaled air quality on their Smartphones and get the secured personalized environment for their comfort and health. The bigger population will have easy access to visualize the quality of air that they are inhaling from their respective locations and time. In general, the control authorities or public get an access to the data streaming about air quality now, the past and predict the future for any location and time [5].

III. METHODOLOGY

The data collected through PAQS sensor devices have the following features:

- ts- Time stamp
- temp- Temperature
- rh- Relative Humidity
- co2- Carbon Dioxide
- co– Carbon Monoxide
- no2- Nitrogen Dioxide
- so2- Sulfur Dioxide
- o3- Ozone
- pm2.5 and pm10- Particulate Matter

Overview of the features selected for data analysis:

Carbon Dioxide (CO2) - This is by far the most emitted form of human caused air pollution. The CO2 increase in earth's atmosphere has been accelerating [2].

Sulfur Dioxide (SO2) - SO2 is produced by volcanoes and in various industrial processes. Coal and petroleum often contain sulfur compounds, and their combustion generates sulfur dioxide. Further oxidation of SO2, usually in the presence of a catalyst such as NO2, forms H2SO4, and thus acid rain. This is one of the causes for concern over the environmental impact of the use of these fuels as power sources [2].

Nitrogen Dioxide (NO2) - Nitrogen dioxide is a chemical compound with the formula NO2. It is expelled from high-temperature combustion, and are also produced during thunderstorms by electric discharge. They can be seen as a brown haze dome above or a plume downwind of cities. It is one of the most prominent air pollutants, this reddish-brown toxic gas has a characteristic sharp, biting odor[2].

Carbon Monoxide (CO) - CO is a colorless, odorless, toxic yet non-irritating gas. It is a product of incomplete combustion of fuel such as natural gas, coal or wood. Vehicular exhaust is a major source of carbon monoxide [2].

Particulate Matter (PM) - Particulates, alternatively referred to as particulate matter (PM), atmospheric particulate matter, or fine particles, are tiny particles of solid or liquid suspended in a gas. In contrast, aerosol refers to combined particles and gas. Some particulates occur naturally, originating from volcanoe2s, dust storms, forest and grassland fires, living vegetation, and sea spray. Human activities, such as the burning of fossil fuels in vehicles, power plants and various industrial processes also generate significant amounts of aerosols. Averaged worldwide, anthropogenic aerosols—those made by human activities— currently account for approximately 10 percent of our atmosphere. Increased levels of fine particles in the air are linked to health hazards such as heart disease, altered lung function and lung cancer [3].

Ozone (O3) – O3 is a key constituent of the troposphere. Ground level ozone (O3) is formed from NOx and VOCs. It is also an important constituent of certain regions of the stratosphere commonly known as the Ozone layer. Photochemical and chemical reactions involving it drive many of the chemical processes that occur in the atmosphere by day and by night. At abnormally high concentrations brought about by human activities (largely the combustion of fossil fuel), it is a pollutant and a constituent of smog [2].

Volatile Organic Compounds (VOC) - VOCs are a well-known outdoor air pollutant. They are categorized as either methane (CH4) or non-methane (NMVOCs). Methane is an extremely efficient greenhouse gas which contributes to enhancing global warming. Other hydrocarbon VOCs are also significant greenhouse gases because of their role in creating ozone and prolonging the life of methane in the atmosphere. This effect varies depending on local air quality. The aromatic NMVOCs benzene, toluene, and xylene are suspected carcinogens and may lead to leukemia with prolonged exposure. 1,3-butadiene is another dangerous compound often associated with industrial use [2].

A. Data Preprocessing:

For analysis of the air pollution data, I had chosen a subset of the huge dataset that was available from the PAQS sensor data. I had chosen the dataset for a period from May 2016 to Feb 2017. Below is a sample (Fig.1) of the dataset after the preprocessing steps such as taking care of missing values and duplicate data and converting date/time values to suitable date/time format:

	temp	rh	co2	co	no2	so2	o3	pm25	pm10
ts									
2016-12-01 00:00:04+05:30	19.99	46.94	331	1.14	0.80	31.53	54.93	0.95	59.84
2016-12-01 00:00:34+05:30	19.99	46.94	346	1.14	0.80	38.97	54.93	1.07	19.95
2016-12-01 00:01:03+05:30	19.99	46.94	341	1.14	0.80	28.57	54.93	0.16	0.00
2016-12-01 00:01:33+05:30	19.99	46.94	333	1.14	0.80	21.11	54.93	0.49	0.00
2016-12-01 00:02:03+05:30	19.99	46.94	332	1.14	0.80	35.27	54.93	1.50	35.07
2016-12-01 00:02:33+05:30	19.99	46.94	339	1.14	0.80	34.31	54.93	1.27	56.95
2016-12-01 00:03:03+05:30	19.99	46.94	337	1.14	0.80	43.00	54.93	1.26	64.99
2016-12-01 00:03:33+05:30	19.99	46.94	349	1.14	0.80	31.77	54.93	0.57	0.00
2016-12-01 00:04:03+05:30	19.99	47.43	357	1.14	0.80	37.03	54.93	0.78	29.92
2016-12-01 00:04:32+05:30	19.99	47.43	339	1.14	0.80	24.03	54.93	1.21	25.10
2016-12-01 00:05:02+05:30	19.99	47.43	345	1.14	0.80	30.84	54.93	0.55	18.02

Fig. 1 (Sample Data)

I started with researching the methods of analysis required for the available data. I wanted to understand the combination of features I have to use to come up with the various effects of air pollution caused by the different pollutants. I wanted to understand the relationship of temperature and relative humidity to the different air pollutants. Another factor to study was the time. I wanted to analyze the levels of pollution at different times of the day and night. As PAQS sensors collect data for every 30 seconds for 24hrs after the data pre-processing I grouped data into separate months, days and hours (Fig.2 and Fig.3).

	temp	rh	co2	co	no2	so2	o3	pm25	pm10
ts									
0	17.714137	55.031705	367.904366	1.14	0.800083	32.462308	57.374179	0.961497	18.349667
1	17.634224	55.736915	376.958592	1.14	0.795611	33.085155	57.384451	1.330621	31.706190
2	17.910352	55.250311	381.836439	1.14	0.755176	32.522526	56.674306	1.931636	73.314099
3	20.136466	48.877048	358.810811	1.14	0.731247	29.033493	56.408524	1.967734	71.645593
4	24.241023	37.525629	324.550107	1.14	0.694861	26.055224	55.372814	1.683987	41.984861
5	29.218691	29.320497	301.361174	1.14	0.671512	18.990000	54.394650	1.262483	38.448284
6	33.126887	21.809091	272.859504	1.14	0.670000	11.713058	53.153581	0.849587	32.500028
7	34.928023	19.970904	281.084746	1.14	0.675000	10.303955	53.025367	0.807684	37.503362
8	33.986309	20.171956	309.721763	1.14	0.703609	14.769284	54.012094	0.695124	30.046612
9	29.733774	23.549532	314.988981	1.14	0.726474	23.517218	54.455758	0.563113	31.800358

Fig. 2 (Data Grouped by Hour)

	temp	rh	co2	co	no2	so2	o3	pm25	pm10
ts									
2016-11-30	19.653313	49.655648	331.602410	1.14	0.800000	34.150316	54.930000	0.536340	17.778087
2016-12-01	24.061642	44.341583	326.483960	1.14	0.763646	27.002497	54.409414	1.320007	49.289000
2016-12-02	23.482388	45.524782	374.870171	1.14	0.764090	27.424246	56.531344	2.008214	31.406906
2016-12-03	21.865950	37.061329	374.797307	1.14	0.765276	29.444523	56.985128	0.831067	44.338653
2016-12-04	18.214835	45.897635	413.471557	1.14	0.749251	31.504057	57.267515	0.506886	21.934551

Fig. 3 (Data Grouped by Date)

This data is taken from the PAQS sensor placed at Kadubeesanahalli location, Bengaluru. Here the time 0 corresponds to 12.00 AM and thereafter it continues through the 24hr period. The PAQS sensor captures time in UNIX format, which I have converted to UTC standard in python.

B. *Data Analysis Visualization Output:*

1. Correlation Matrix:

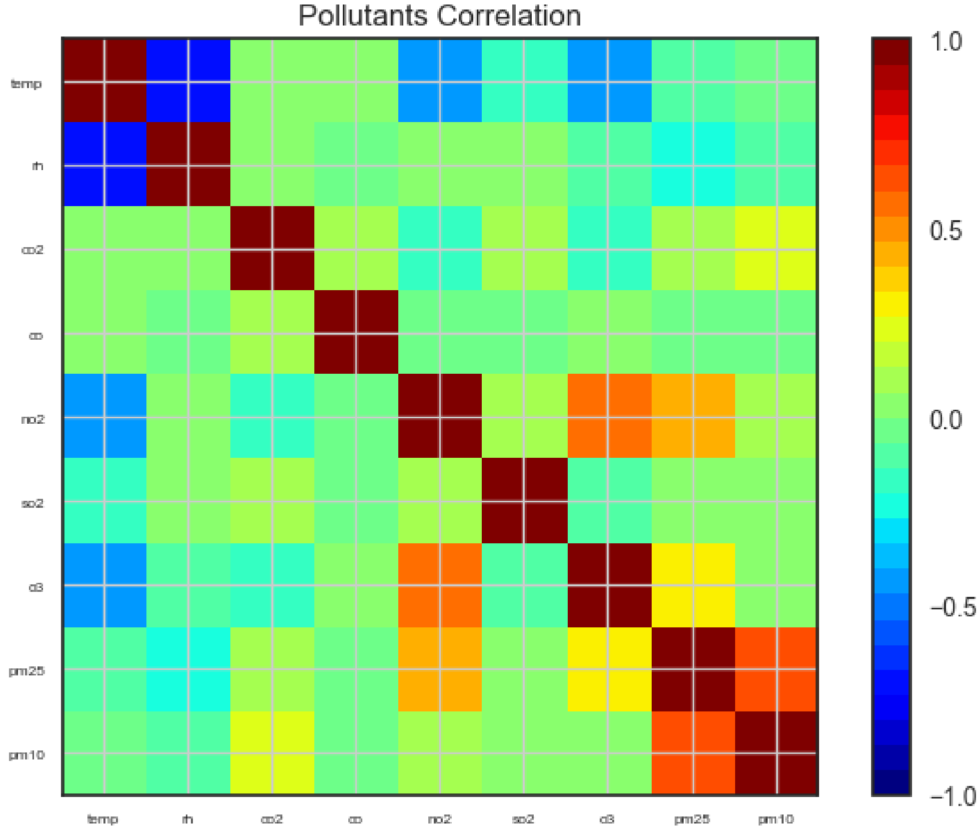

Fig. 4 (Pollutants Correlation Matrix)

From the correlation matrix (Fig.4) the following observations can be made:

- Strong negative correlation between temperature and relative humidity.
- The negative correlation between temperature and no2, temperature and o3.
- Strong positive correlation between pm2.5 and pm10.
- A positive correlation between no2 and o3, no2 and pm2.5.

2. *Trend/Seasonal Analysis:*

Below is the trend analysis (Fig.5) for all the features on an average time basis for the period from May 2016 to Feb 2017:

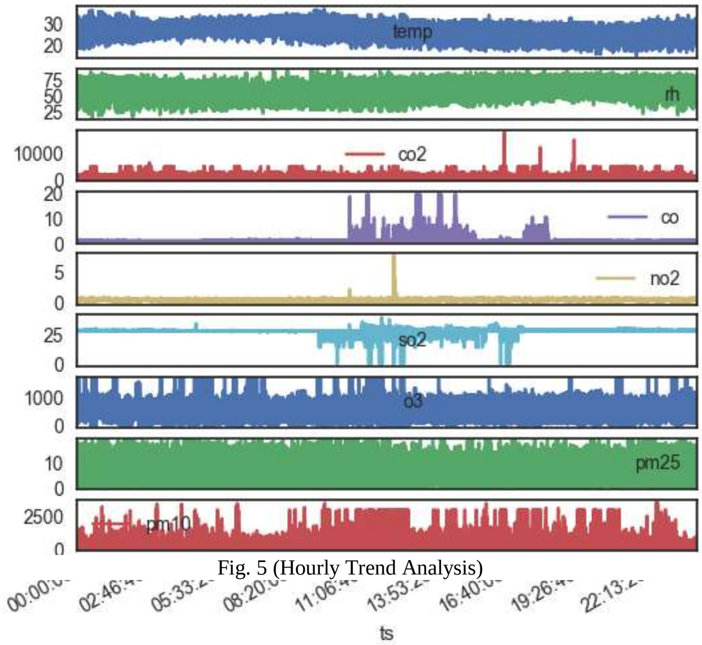

Fig. 5 (Hourly Trend Analysis)

Below is the trend analysis

(Fig. 6) on monthly basis for these 10 months:

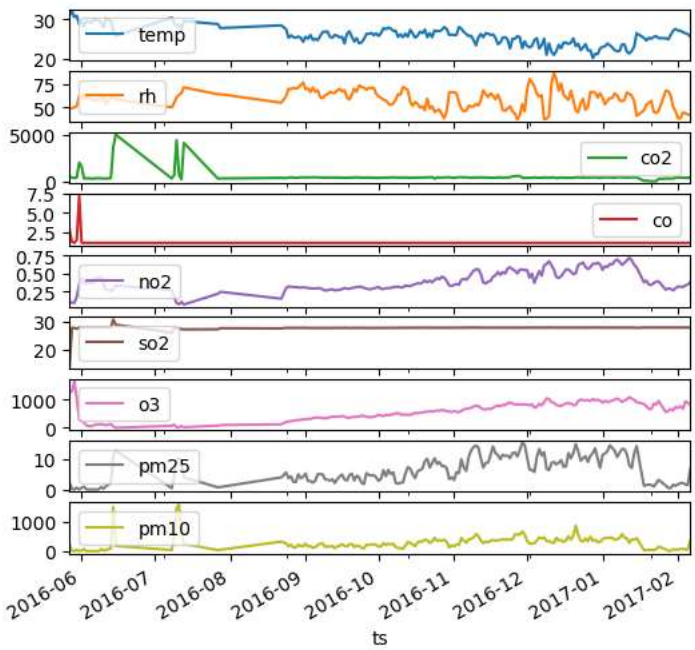

Fig. 6 (Monthly Seasonal Analysis)

Following observations can be made from the trend/seasonal analysis outputs:

- The key pollutants which are pm2.5, pm10 and o3 are on a high during peak traffic hours during the day time. Variation in the levels can be seen especially in pm10 during early morning and late evenings.
- The level of most pollutants is seen increasing linearly from the month June to February i.e. during the monsoons the pollutant levels are low and then they gradually increase during the winter and spring seasons. Later again the level decreases during summer.

3. *Box Plots:*

	temp	rh	co2	co	no2	so2	o3	pm25	pm10
count	24483.000000	24483.000000	24483.000000	2.448300e+04	24483.000000	24483.000000	24483.000000	24483.000000	24483.000000
mean	25.417131	55.405494	257.783850	1.140000e+00	0.373556	27.775716	757.201231	3.246309	117.333310
std	2.639429	12.636025	287.916628	3.692677e-13	0.140546	0.042795	225.320536	4.904295	266.420575
min	15.160000	21.640000	0.000000	1.140000e+00	0.000000	27.330000	27.470000	0.000000	0.000000
25%	24.820000	44.510000	0.000000	1.140000e+00	0.280000	27.760000	586.610000	0.560000	0.270000
50%	25.500000	55.760000	295.000000	1.140000e+00	0.310000	27.780000	769.070000	1.120000	27.830000
75%	26.870000	66.235000	355.000000	1.140000e+00	0.390000	27.800000	931.900000	2.500000	106.500000
max	32.390000	91.050000	24247.000000	1.140000e+00	0.800000	29.360000	1304.670000	18.780000	3875.750000

Fig. 7 (Box Plot values - Jan 2017)

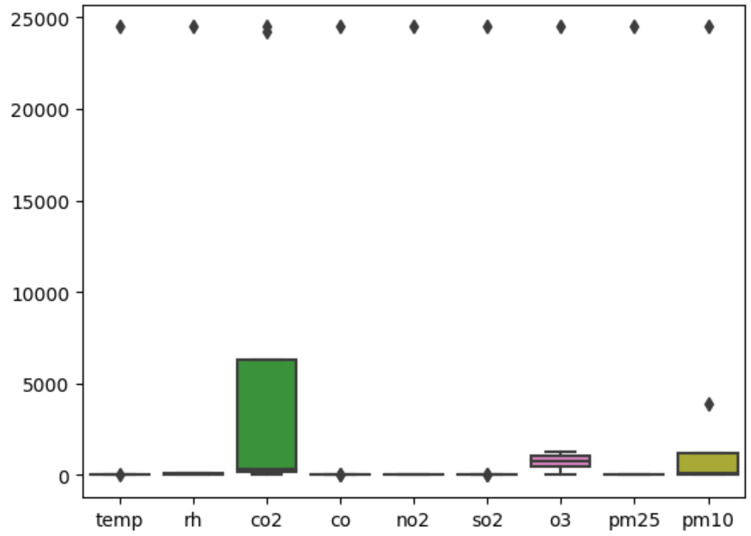

Fig. 8 (Box Plot - Jan 2017)

	temp	rh	co2	co	no2	so2	o3	pm25	pm10
count	30439.000000	30439.000000	30439.000000	3.043900e+04	30439.000000	30439.000000	30439.000000	30439.000000	30439.000000
mean	29.228526	58.935636	726.463944	1.140000e+00	0.369649	27.982082	98.812239	1.364968	128.333677
std	1.537116	4.178512	1271.795736	4.107893e-13	0.083307	1.163415	102.513064	3.700378	505.630197
min	0.000000	0.000000	136.000000	1.140000e+00	0.060000	27.580000	0.000000	0.000000	0.000000
25%	28.260000	56.740000	274.000000	1.140000e+00	0.340000	27.730000	54.930000	0.000000	0.000000
50%	28.940000	59.680000	311.000000	1.140000e+00	0.370000	27.760000	72.590000	0.000000	0.000000
75%	30.330000	61.150000	355.000000	1.140000e+00	0.400000	27.780000	117.710000	0.330000	0.000000
max	33.080000	74.390000	5000.000000	1.140000e+00	0.690000	45.160000	1002.530000	14.090000	2906.650000

Fig. 9 (Box Plot values - June 2016)

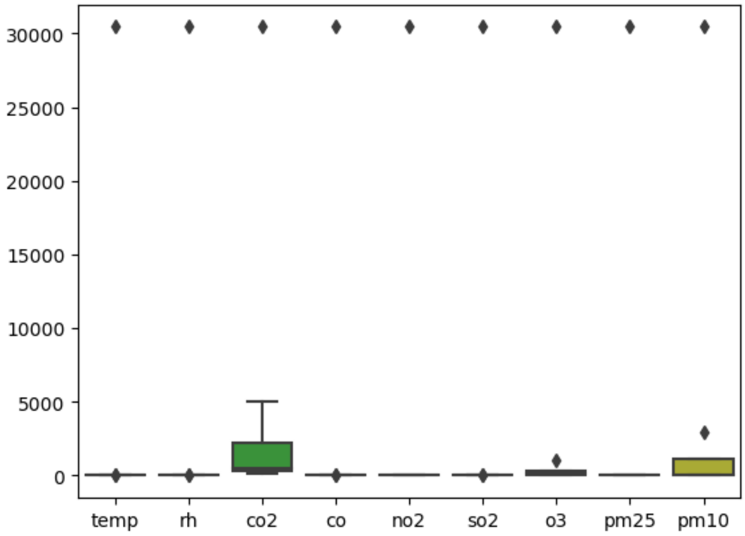

Fig. 10 (Box Plot - June 2016)

From the box plots following observations can be made:

- Overall the pollutant levels are high in January when compared to June. This might be because Bangalore usually gets some rainfall in June which in turn reduces the relative humidity.
- There are few outliers for both Jan and June values.

IV. AIR QUALITY INDEX (AQI)

An AQI is defined as an overall scheme that transforms weighted values of individual air pollution related parameters (SO2, CO, etc.) into a single number or set of numbers. AQI is useful for the general public to know air quality in a simplified way. Air quality standards are the basic foundation that provides a legal framework for air pollution control. An air quality standard is a description of a level of air quality that is adopted by a regulatory authority as enforceable [4].

Method to Compute AQI

Many AQI computations and depend on the measurement of pollutant concentrations as prescribed.

The formula for computing AQI for a specific pollutant is

$\mathbf{I_p = [\{(I_{HI} - I_{LO})/ (B_{HI} - B_{LO})\} * (C_p - B_{LO})] + I_{LO}}$

B_{HI} = Breakpoint concentration greater or equal to given conc.

B_{LO} = Breakpoint concentration smaller or equal to given conc.

I_{HI} = AQI value corresponding to B_{HI}

I_{LO} = AQI value corresponding to B_{LO}

C_p = Pollutant value

The index is finally arrived at by using the value of AQI for the maximally polluting component.

AQI = Max (I_p) (where; p= 1,2,...,n; denotes n pollutants)

Indian Air Quality Index

India Air Quality standards are defined for the pollutants with short-term standards (1hr, 8hrs, 24hrs) shown in the Table. 1. All quantities are measured in Micrograms/m^3 except for CO.

Table 1. Indian National Air Quality Standards

Pollutant	SO_2	NO_2	$PM_{2.5}$	PM10	O3		CO (mg/m^3)		Pb	NH_3
Averaging time (hr)	24	24	24	24	1	8	1	8	24	24
Standard	80	80	60	100	180	100	4	2	1	400

These pollutants, except for CO and O3, have annual standards as well. In addition, Pb (lead), As (Arsenic), BaP (benzopyrene), C6H6 (Benzene) and Nickel are also identified as potentially harmful pollutants, but only have annual standards.

Table 2. Break Points for computing AQI in India.

AQI Category (Range)	PM10 24-hr	$PM_{2.5}$ 24-hr	NO_2 24-hr	O3 8-hr	CO 8-hr (mg/ m^3)	SO_2 24-hr	NH_3 24-hr	Pb 24-hr
Good (0-50)	0-50	0-30	0-40	0-50	0-1.0	0-40	0-200	0-0.5
Satisfactory (51-100)	51-100	31-60	41-80	51-100	1.1-2.0	41-80	201-400	0.5 –1.0
Moderately polluted (101-200)	101-250	61-90	81-180	101-168	2.1- 10	81-380	401-800	1.1-2.0
Poor (201-300)	251-350	91-120	181-280	169-208	10-17	381-800	801-1200	2.1-3.0
Very poor (301-400)	351-430	121-250	281-400	209-748*	17-34	801-1600	1200-1800	3.1-3.5
Severe (401-500)	430 +	250+	400+	748+*	34+	1600+	1800+	3.5+

Use the colors indicated in Table. 2 for different categories of severity of air pollution.

Table 3. Health Advisory

AQI	Associated Health Impacts
Good(0–50))	Minimal Impact
Satisfactory (51–100)	May cause minor breathing discomfort to sensitive people
Moderately polluted (101–200)	May cause breathing discomfort to the people with lung disease such as asthma and discomfort to people with heart disease, children, and older adults
Poor (201–300)	May cause breathing discomfort to people on prolonged exposure and discomfort to people with heart disease
Very Poor (301–400)	May cause respiratory illness to the people on prolonged exposure. Effect may be more pronounced in people with lung and heart diseases
Severe (401-500)	May cause respiratory effects even on healthy people and serious health impacts on people with lung/heart diseases. The health impacts may be experienced even during light physical activity

Below are few AQI samples calculated on an hourly basis for PAQS data based on the above-given formula for computing AQI.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
pm10	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	139.0	142.0
pm25	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
no2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
so2	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0

Fig. 11 (AQI - Oct 2, 2016)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
pm10	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	222.0	189.0
pm25	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	7.0
no2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
so2	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0

Fig. 12 (AQI - Oct 4, 2016)

AQI is calculated for these features for the 24hrs range. The above readings (Fig.11, Fig.12, Fig.13) shows AQI for each pollutant. Overall AQI for the day will be the highest AQI amongst all the AQIs. For example for Oct 27 overall AQI will be 322. As per these readings, the AQI for PM10 is very high and if all the days AQI readings are observed then AQI for PM10 is high. AQI for PM25, NO2, and SO2 are in the minimal impact range.

For 2010, the key findings of India's central pollution control board are [4]:

- Most Indian cities continue to violate India's and world air quality PM10 targets. Respirable particulate matter pollution remains a key challenge for India.
- Most Indian cities greatly exceed acceptable levels of suspended particulate matter. This may be because of biomass burning, vehicles, power plant emissions, and industrial sources.
- The Indian air quality monitoring stations reported lower levels of PM10 and suspended particulate matter during monsoon months possibly due to wet deposition and air scrubbing by rainfall.
- Higher levels of particulates were observed during winter months possibly due to lower mixing heights and more calm conditions. In other words, India's air quality worsens in winter months and improves with the onset of monsoon season.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
pm10	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	332.0	293.0
pm25	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	15.0	13.0
no2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
so2	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0

Fig. 13 (AQI - Oct 27, 2016)

The observation about PM level being low in monsoon and high comparatively in winter season is demonstrated in the PAQS sensor data as well.

For June 2016:
PM2.5 AQI is 5, which is good.
PM10 AQI is 87, which is satisfactory.

For November 2016:
PM2.5 AQI is 41, which is satisfactory.
PM10 AQI is 228, which is bad.

RESULT

Based on the various data visualization and analysis done it can be evaluated that most of the pollutants vary within acceptable limits and do not have significant correlation amongst other pollutants. However PM10 value is quite high for most of the months. As these readings are for the Kadubeesanahalli sensor the reason for this excess PM10 level might be due to high vehicular density in the vicinity. Along with that other factors such as wind speed, road repairs, building construction that result in high dust particles might also have some impact on the variation in pollutant levels.

CONCLUSION

Based on data study and research carried out for the air pollutant dataset collected by PAQS sensor device it can be inferred that the pollutant levels of some harmful particulate matters such as PM10 are quite high in the air. One of the key sources for this seems to be the high vehicle density in the location where this data was collected. As the pollutants are mostly varying in a constant manner throughout the day the approximate pollutant levels can be predicted for the next day for a particular temperature and relative humidity factor. Also based on the observation it can be predicted that the pollutant levels vary based on the vehicle density i.e. during peak hours it is quite high. Also, air pollutants are influenced by the seasonal changes. In monsoon, the pollutant levels are low when compared to winter and spring seasons where it is high and again in summer the pollutant levels go down.

REFERENCES

[1] "*Air Pollution Causes, Effects, and Solutions*". National Geographic. 9 October 2016.

[2] "*Environmental Pollution*". Theenvironmentalblog.org. 2011-12-16. Retrieved 2012-12-11.

[3] https://www.marlborough.govt.nz/environment/air-quality/smoke-and-smog/health-effects-of-pm10.

[4] "*Ambient Air Quality Data at various locations*". Central Pollution Control Board. 2015-16.

[5] http://www.paqs.biz/.

