DATA DESCRIPTION DOCUMENTATION

Group Name : Chhoti Ki Toli

1.Dataset Description:

The given data is related to air pollution. This dataset contains different pollutants that affect the air pollution of a city. The places under consideration are 8 different areas of a particular city named Vadodara of an Indian state, Gujarat. The pollutants include NO_2, SO_2, CO,etc,. The data consists of different values for each parameter at a given time and at a particular place.

2.Name of Data file:

Environmental Report.

3.Language:

English.

4.Owner of Data Source:

State Government of Gujarat, India.

5.File Size:

39MB
Number of rows – 3,27,465
Number of columns – 18

6.Methodology for data collection:

The data was collected using sensors. The more common way of getting data out of smart sensors is to use a bridging device known as a gateway in each room. A gateway receives data from the sensors and makes it usable. Data is transmitted from the sensors to the gateway wirelessly. But in our case data is transferred through optical fiber on the server. The sensors used here are Environmental sensors of PAQS.

7.Purpose of Data Collection:

The sole purpose of collecting the data was to analyze where the emission of CO_2 and SO_2 are more to take necessary actions in the respective areas.

8.Data Format:

The data provided here is in a structured format. The file is in Microsoft Excel spreadsheet format. Excel format supports more complex data. Excel also enables users to identify trends and organize and sort data into meaningful categories.

9.Timing of Data Collection:

The data was collected between the dates Dec 31st,2018 to Oct 24,2020. Through these dates the data is collected at different time stamps throughout the day.

10.Data Form:

The data was collected every day at an interval of 15 minutes at every station. An example of the same is given below.

Startdatetime	Aqi Param	Aqi Value	Co	Co 2	Light	Name	Noise	No 2	O 3	Pm 10	Pm 25	Remark	Rh	So 2	Temp	Ts	Uva	U
Dec 31, 2018 6:30:00 PM			0.65	507.67	3	Atladara Pumping Station	50.26	20.46	65.29	228.09	8.33		57.23	54.71	18.61		0.1	
Dec 31, 2018 6:30:00 PM			0.42	366.68	2	Channi Toll Naka	55.45	30.01	76.2	187.44	6.45		62.77	18.72	17.23		0.04	0
Dec 31, 2018 6:30:00 PM			0.58	572.19	2	Jambuva Police Check Post	53.1	20.66	66.01	302.71	9.1		64.09	89.78	17.23		0	
Dec 31, 2018 6:30:00 PM			1.55	444.47	3	Makarpura GIDC	46.36	13.52	51.74	358.88	11.87		60.38	91.43	17.92		0.06	0
Dec 31, 2018 6:45:00 PM			0.67	514.87	3	Atladara Pumping Station	40.82	20.13	64.19	221.23	8.57		57.51	53.92	18.61		0.1	
Dec 31, 2018 6:45:00 PM			0.42	371.6	2	Channi Toll Naka	50.6	29.05	75.29	156.48	6		63.35	20.4	16.78		0.04	0
Dec 31, 2018 6:45:00 PM			0.52	509.23	2	Jambuva Police Check Post	48.12	20.66	66	215.44	7.34		59.72	35.84	17.9		0	
Dec 31, 2018 6:45:00 PM			1.05	423.63	3	Makarpura GIDC	42.72	13.28	51.24	307.58	11.84		58.81	81.62	17.96		0.06	0
Dec 31, 2018 7:00:00 PM			0.6	514.03	3	Atladara Pumping Station	42.45	20.37	64.96	190.46	7.71		57.52	44.64	18.61		0.1	
Dec 31, 2018 7:00:00 PM			0.41	370.5	2	Channi Toll Naka	55.24	29.63	75.83	181.46	5.85		63.6	17.39	16.55		0.04	0
Dec 31, 2018 7:00:00 PM			0.53	498.52	2	Jambuva Police Check Post	48.53	20.95	67.01	233.96	7.65		59.05	42.45	17.92		0	
Dec 31, 2018 7:00:00 PM			0.57	415.53	3	Makarpura GIDC	40.43	15.33	54.08	158.99	5.85		57.99	65.42	18.61		0.06	0
Dec 31, 2018 7:15:00 PM			0.51	502.33	3	Atladara Pumping Station	40.83	21.25	67.93	149.16	6.23		56.95	37.91	18.61		0.1	
Dec 31, 2018 7:15:00 PM			0.41	371.94	2	Channi Toll Naka	49.13	29.87	76.07	148.7	5.2		64.46	15.49	16.55		0.04	0

11.Data Confidentiality and Access:

The given data is used by the State Government of Gujarat, India. Therefore, this makes the data very confidential and reduces its access to a minimum. The data can only be retrieved upon special request and permission from the respective authorities.

12.Variables:

1. StartDate Time-

It is the information about the day and date at which the readings were taken and at what exact time were they taken.

Datatype – Date/Time

Numerical (Date – discrete , Time – continuous)

2. Aqi Parameter-

It is the parameter or the pollutant on which the Air Quality Index has been calculated.

Datatype – Text

Categorical (nominal)

(Since it is qualitative data mathematical operations cannot be performed)

3. Aqi Value-

An air quality index (AQI) is used by government agencies to communicate to public how polluted the air currently is or how polluted it is forecast to become.
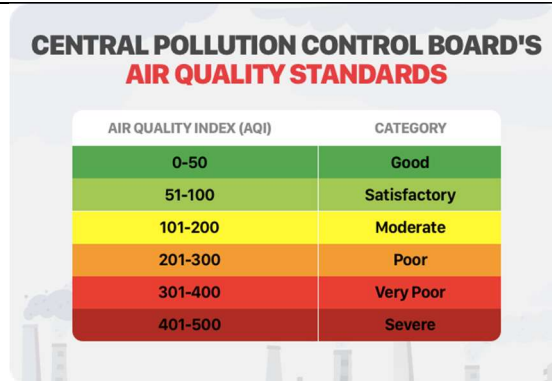
All the eight pollutants may not be monitored at all the locations. Overall AQI is calculated only if data are available for minimum three pollutants out of which one should necessarily be either PM2.5 or PM10. Else, data are considered insufficient for calculating AQI. Similarly, a minimum of 16 hours' data is considered necessary for calculating subindex. The sub-indices for monitored pollutants are calculated and disseminated, even if data are inadequate for determining AQI. The Individual pollutant-wise sub-index will provide air quality status for that pollutant.

Datatype – Whole number
             Numerical (discrete)

Unit- ug/ m^3

(Other mathematical operations cannot be performed on the raw data because some of the values are null values which is a textual datatype)

CENTRAL POLLUTION CONTROL BOARD'S
AIR QUALITY STANDARDS

AIR QUALITY INDEX (AQI)	CATEGORY
0-50	Good
51-100	Satisfactory
101-200	Moderate
201-300	Poor
301-400	Very Poor
401-500	Severe

4. Co-

It is a colorless gas, released from automobile emissions, fires, industrial processes, gas stoves, kitchen chimneys, generators, wood-burning smoking, etc. into the atmosphere.

Safe range - 0-4 mg/m³ (1-hour)

Unit- mg/m³

Datatype – Decimal

Numerical (continuous)

Max – 4.54

Min – 0

Mean – 0.456590

Median – 0.38

Range – 4.54

Standard Deviation - 0.468823

5. Co2-

It is a colorless gas having a faint sharp odour and a sour taste. It is one of the most important greenhouse gases linked to global warming.

Safe range – 5000ppm (permissible exposure limit for daily workplace exposures.)

Datatype – Decimal

Numerical (continuous)

Unit – ppm

Max – 1932.5

Min – 0

Mean – 388.418927

Median – 394.18

Range – 1932.5

Standard Deviation – 138.046803

6. Light –

Luminous intensity is the amount of light that falls on a surface across a certain square foot or square meter.

Range- 500 - 1000 lux

Datatype – Decimal
                Numerical (continuous)

Unit – lux

Max – 9691.77

Min – 0

Mean – 1451.929217

Median – 14.81

Range – 9691.77

Standard Deviation – 2232.1730571465368

7. Area Name-

8 different areas of a particular city namely-

Mangal Bazar -- Jan 4, 2019 7:30:00 PM to October 24,2020 12:27:46

Transpek Area -- Dec 31, 2018 7:30:00 PM to October 24,2020 12:28:31 PM

Altadara Pumping Station -- Dec 31,2018 6:30:00 PM to October 24, 2020 12:27:23 PM

Makarpura GIDC -- Dec 31, 2018 6:30:00 PM to October 24,2020 12:27:00 PM

Jambuva Police Check Post -- Dec 31, 2018 6:30:00 PM to October 24,2020 12:28:55 PM

Gorwa Pumping Station -- Jan 5,2019 1:15:00 AM to October 24,2020 12:20:42 PM

Bapod (Ward 9) Office -- Jan 1,2019 3:15:00 PM to October 24,2020 12:28:09 PM

Channi Toll Naka -- Dec 31,2018 6:30:00 PM to October 24, 2020 12:29:18 PM

Datatype – Text

         Categorical (nominal)

(Since it is qualitative data mathematical operations cannot be performed)

8. Noise

Noise is unwanted sound. Noise pollution, also known as environmental noise or sound pollution, is the propagation of noise with ranging impacts on the activity of human or animal life, most of them are harmful to a degree.

Safe range - below 70 dBA over 24-hours (75 dBA over 8-hours)

Datatype – Decimal

         Numerical (continuous)

Unit – Decibel

Max – 82.54

Min – 0

Mean – 49.259011

Median – 49.18

Range – 82.54

Standard Deviation - 12.270846

9. No2-

Nitrogen Dioxide (NO_2) is one of a group of highly reactive gases known as oxides of nitrogen or nitrogen oxides (NO_x). Nitrogen dioxide is a reddish-brown gas with a pungent. It highly reactive gas present in the atmosphere.

Safe range - 0-80 ug/m^3 (24 hours)

Datatype – Decimal

         Numerical (continuous)

Unit – ug/m^3

Max – 156.88

Min – 0

Mean – 25.471575

Median – 14.04

Range – 156.88

Standard Deviation - 28.495779

10.   O3-

Ozone is composed of three oxygen atoms. It forms the protective layer which prevents entry of harmful ultraviolet radiation into the earth. The ground ozone is very harmful to human beings and the environment.

Safe range - 0-100 ug/m³ (8 hours)

Datatype – Decimal

Numerical (continuous)

Unit – ug/m³

Max – 111.43

Min – 0

Mean – 81.561075

Median – 85.43

Range – 111.43

Standard Deviation - 19.552614

11.   Pm10-

A mixture of particles with liquid droplets in the air forms particulate matter. PM 10 are particles that have a size of less than or equal to 10 microns.

Safe range - PM10 (24 hours) are 0-100 ug/m³

Datatype – Decimal

Numerical (continuous)

Unit- ug/m³

Max – 5092.27

Min – 0

Mean – 157.939982

Median – 129.63

Range – 5092.27

Standard Deviation - 145.417867

12.   Pm2.5 –

A mixture of particles with liquid droplets in the air forms particulate matter. PM2.5 are ultra-fine particles having a size of less than or equal to 2.5 microns.

Safe range - PM 2.5 (24 hours) is 0-60 ug/m³

Datatype – Decimal

Numerical (continuous)

Unit- ug/m³

Max – 649.11

Min – 0

Mean – 37.278513

Median – 28.44

Range – 649.11

Variation - 38.688701

13.   Remark

It is the remark about the pollution at that place stating whether it is good or bad. The types are:

Satisfactory

Good

Sensor Down
Moderately Polluted
Poor
Very Poor
Severe
Unhealthy for sensitive groups
Datatype – Text
                Categorical (ordinal)

14.   Rh-

The term relative humidity (RH) expresses the relationship between the moisture content of air at a certain temperature and the moisture content of moisture-saturated air at the same temperature.
Safe range – 30% to 50%
Datatype – Decimal
                Numerical (continuous)
Max – 110.66
Min – 0
Mean – 52.362205
Median – 68.63
Range – 110.66
Variation - 26.282335

15.   So2-

Sulfur dioxide is a colorless gas with a burnt odor and the chemical formula SO_2. The gas is acidic & corrosive in nature and can react in the atmosphere with other compounds to form sulfuric acid and other oxides of sulfur.

Safe range - 0-80 ug/m^3 (24 hours)

Datatype – Decimal

           Numerical (continuous)

Unit- ug/m^3

Max – 785.83

Min – 0

Mean – 12.343422

Median – 2.62

Range – 785.83

Variation - 25.932111

16.   Temperature-

Temperature is the measure of hotness or coldness expressed in terms of any of several scales, including Fahrenheit and Celsius.

Safe range - 15°C TO 42.3^{o}C

Datatype – Decimal

           Numerical (continuous)

Unit- ^{o}C

Max – 49.28

Min – 0

Mean – 27.805033

Median – 28.42

Range – 49.28

Standard Deviation - 7.817154

17.   Ts-

Total suspended particulate (TSP) refers to the totality of small solid matter released, documented and/or otherwise observed in the atmosphere. Total suspended particulates are considered to be a

primary contributor to air pollution, smog formation and environmental contamination. TSP consist of two main particle sizes: PM_{10} and $PM_{2.5}$.

18. Uva-

UVA rays also play a role in skin cancer formation. In addition, the UVA rays penetrate more deeply into the skin and play a greater role in premature skin aging changes including wrinkle formation.

Safe range – 315-400 nm

Datatype – Decimal

Numerical (continuous)

Unit- nm

Max – 4.04

Min – 0

Mean – 0.091696

Median – 0

Range – 4.04

Variation - 0.254263

19. Uvb-

UVB rays are responsible for producing sunburn. The UVB rays also play the greatest role in causing skin cancers, including the deadly black mole form of skin cancer (malignant melanoma).

Safe range – 280-315 nm

Datatype – Decimal

Numerical (continuous)

Unit- nm

Max – 4.04

Min – 0

Mean – 0. 091696

Median – 0

Range – 4.04

Variation - 0.254263

Table 3.11 Breakpoints for AQI Scale 0-500 (units: µg/m³ unless mentioned otherwise)

AQI Category (Range)	PM₁₀ 24-hr	PM₂.₅ 24-hr	NO₂ 24-hr	O₃ 8-hr	CO 8-hr (mg/m³)	SO₂ 24-hr	NH₃ 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	0.6 –1.0
Moderate (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	10.1-17	381-800	801-1200	2.1-3.0
Very poor (301-400)	351-430	121-250	281-400	209-748*	17.1-34	801-1600	1201-1800	3.1-3.5
Severe (401-500)	430 +	250+	400+	748+*	34+	1600+	1800+	3.5+

*One hourly monitoring (for mathematical calculation only)