International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 5 ǁ 2022 ǁ PP. 26-34

Research on Air Quality Index: A Comparative Study to Determine Air Quality

Abhinav Saxena[1], Shweta Sunil Yadav[2], Anandita Bajpai[3]
[1]*Student Information technology KIET Group of Institutions Uttar Pradesh, India*
[2]*Student Computer Science & Information Technology KIET Group of Institutions Uttar Pradesh, India*
[3]*Student Information technology KIET Group of Institutions Uttar Pradesh, India*

Nitin Kumar[4], Ambrish Gangal[5]
[4]*Professor Information Technology KIET Group of Institutions Uttar Pradesh, India*
[5]*Professor Computer Science & Information Technology KIET Group of Institutions Uttar Pradesh, India*

ABSTRACT
The Air Quality Index is a technique for determining the current state of air quality. The Air Quality Index (AQI), which is based on the synergistic four pollutants (PM10, PM2.5, SO2, and NO2), was used to compare the current ambient air quality in the research region. ***Why is air quality so important?*** *is the question that arises and answer to that is:-The quality of your local air has an impact on how you live and breathe. It can change from day to day, or even hour to hour, much like the weather. The US Environmental Protection Agency (EPA) and your local air quality agency have been attempting to make outdoor air quality information as accessible and understandable as weather forecasts.*
The Air Quality Index, or AQI is a crucial instrument in this effort.

Date of Submission: 24-04-2022 Date of acceptance: 06-05-2022

I. INTRODUCTION

The Air Quality Index (AQI) is a daily air quality index. It shows you how clean or dirty your air is, as well as any potential health consequences. The AQI measures the health impacts you might have a few hours or days after breathing polluted air. Ground-level ozone, particle pollution, carbon monoxide, and Sulphur dioxide are the four principal air pollutants regulated by the Clean Air Act, and the AQI is determined for each of them. To safeguard public health, the EPA has developed national air quality guidelines for each of these contaminants. The EPA is currently evaluating the national nitrogen dioxide air quality standard. If the standard is changed, the AQI will also be changed. The higher the AQI, the less clean the air is and the greater the risk to one's health. The AQI is used to link its value to a specific health risk. The six AQI categories and the various levels of health concern are shown in Table 1.

Table 1: The Six Categories of AQI

AQI Values	Level of Health Concern	Colour
0 - 50	Good	Green
51 – 100	Moderate	Yellow
101– 150	Unhealthy for Sensitive groups	Orange
151 - 200	Unhealthy	Red
201 – 300	Very Unhealthy	Purple
301 – 500	Hazardous	Maroon

Each level of health concern relates to distinct category:

Excellent Your community's AQI value is between Good and Excellent. The AQI for your neighborhood is between 0 and 50. The air quality is good, and there is little or no harm to one's health.

Moderate. The AQI ranges from 51 to 100.The air quality is satisfactory; but, for a limited number of people, pollution in the range may constitute a moderate health Respiratory symptom may occur in people are highly sensitive to ozone or particle pollution.
Harmful to sensitive Groups. Members of sensitive populations may experience health impacts when AQI levels are between 101 and 150, although the public is unlikely to be harmed.

Ozone. People with lung problems, children the elderly, and smokers are all affected by ozone. Outdoor activity makes people more sensitive, putting them at a higher risk.
Particle Pollution. People with heart or lung problems, older adults, and children are all considered vulnerable to particle pollution and so at a higher risk.
Unhealthy. AQI levels are between 151 and 200, everyone may start to feel the impacts. Members of vulnerable groups may suffer more serious health consequences.
Dangerous. Health alerts are issued when the AQI exceeds 300.

II. CONSTITUENTS OF AIR

Particle pollution (sometimes called "**particulate matter"**) is a mix of solids and liquid droplets. Some particles are discharged directly, while others develop in the atmosphere as contaminants from diverse sources react. During catastrophes such as forest fires, particle pollution levels can be extremely harmful and even dangerous. Indoor particle levels can rise, especially if outdoor particle levels are high.
Particles are available in a variety of sizes. Those with a diameter of less than 10 micrometers (less than the width of a single human hair) are so minuscule that they can enter the lungs and cause major health concerns. Particles that are very fine. **The smallest particles (those with a diameter of 2.5 micrometers or less) are referred to as "fine" particles.** Only an electron microscope can detect these particles because they are so tiny. Motor cars, power plants, domestic wood burning, forest fires, agricultural burning, various industrial activities, and other combustion processes are all major sources of fine particles. Particles that are coarse. The term "coarse" refers to particles with a diameter of 2.5 to 10 micrometers. Crushing and grinding processes, as well as dust stirred up by vehicles travelling on highways, are all sources of coarse particles.

When sulfur-containing fuels like coal and oil are burned, sulfur dioxide, a colorless, reactive gas, is generated. The highest levels of **sulfur dioxide** are complexes of a huge scale Power is one of the most important sources. . Because of its impact on humans and the environment, ozone is a dangerous principal ingredient in "smog." Learn more about the sources of air pollution. What cause ground level ozone? Tropospheric ozone, also known as ground-level ozone, is produced by chemical interactions between nitrogen oxides (NOx) and volatile organic molecules (VOC). When pollutants from automobiles, power plants, industrial boilers, refine rises, chemical plants, and other sources react chemically in the presence of sunshine, this occurs. **Carbon monoxide** is a gas that has no odor and is colorless. When the carbon in fuels does not entirely burn, it produces Vehicle exhaust accounts for around 75% of all carbon monoxide emissions in the United States, and up to 95% in cities.

Ozone is a gas made up of three oxygen atoms. Ozone is found in both the high and lower atmospheres of the Earth. Depending on where it is found, ozone can be beneficial or harmful. Good ozone, also known as stratospheric ozone, is found in the upper atmosphere, where it forms a protective layer that protects us from the sun's harmful UV rays. Man-made chemicals have partially destroyed this beneficial ozone, resulting in a so-called "hole in the ozone." The good news is that the gap is closing.

III. EFFECTS OF WEATHER ON AIR QUALITY

3.1 HOT WEATHER

When the temperature is hot, key pollution sources tend to grow. Heat and sunlight can turn main air pollutants into secondary pollutants that are even more dangerous.

3.2 FORMATION OF SMOG

The dangers of primary and secondary contaminants are different. The first are those that are emitted directly from a source, which might be natural (for example, volcanic eruptions or fires) or anthropogenic (for example, industrial processes) (carbon monoxide from vehicles).

Secondary pollutants, on the other hand, are not immediately released into the atmosphere. The interactions between primary emissions in the atmosphere are the source of it.

High atmospheric pressure causes a stagnant layer of air above ground level during heat waves. When this occurs in cities, air pollutants are trapped, and the density of pollutants rises, and these Heat waves can result in poor air quality. During a heat wave, the intense heat and stagnant air increase the amount of ozone and particulate pollution. Drought conditions, in which soils are extremely dry, can also occur during a heat wave. Forest fires are more prevalent during droughts. Carbon monoxide and particulate pollution are released into the atmosphere by fires.

3.3 COLD WEATHER

Exhaust from vehicles, chimneys, and smokestacks is more visible when the weather is cold. Is this a sign of more pollutants in the air, or simply that the warm vapor exhaust is more visible? In most cases, both are correct. During the frigid winter months, particulate matter and carbon monoxide pollution from wood burning increase. Idling cars to defrost or keep them warm also contributes to air pollution.

3.4 CLOUDINESS

Cool days and low mixing heights are associated with cloudy sky, but sunny days have the reverse effect. The majority of the pollution outbreaks in the NCR this month corresponded with cloudy skies. Low clouds raise humidity, which can cause fog to occur.

3.5 RAINY WEATHER/HUMIDITY

Rain sweeps away contaminants and cleans the air swiftly. Wet spells, on the other hand, leave behind high humidity, which might result in fog in calm breezes. Fog contributes to pollution because water droplets combine with contaminants as secondary particles. On November 2, extremely small rain (0.5mm) in Delhi resulted in dense pollution the following day, raising the AQI to 494. It was the city's second worst pollution incident since 2015.

4. RELATED WORK

Data: We can deal with the problem by confining our research to the most relevant data for the same. In our case we found AQI can serve the best results.

Analysis: With proper tools and accurate data, comes refined results. Here we will work on data related to AQI and its various dependent parameters for analysis.

Flow:

Data Collection--------->Data Cleaning--------->Data Analysis

Data Collection

Data collection the is process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.

Data Cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

Data Analysis

Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense, and recap, and evaluate data.

Data is taken from

Kaggle dataset, Air Quality Historical Data Platform, Metrological Data

Libraries used

NumPy

NumPy is **a Python library used for working with arrays**. It also has functions for working in domain of linear algebra, Fourier transform, and matrices.

Pandas

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python.

Matplotlib

matplotlib. pyplot is a collection of functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

Seaborn

Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures. Seaborn helps you explore and understand your data.

```
df.describe()
```

	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	
count	24933.000000	18391.000000	25949.000000	25946.000000	25346.000000	19203.000000	27472.000000	25677.000000	25509.000000	23908.000000	21490.000000	114
mean	67.450578	118.127103	17.574730	28.560659	32.309123	23.483476	2.248598	14.531977	34.491430	3.280840	8.700972	
std	64.661449	90.605110	22.785846	24.474746	31.646011	25.684275	6.962884	18.133775	21.694928	15.811136	19.969164	
min	0.040000	0.010000	0.020000	0.010000	0.000000	0.010000	0.000000	0.010000	0.010000	0.000000	0.000000	
25%	28.820000	56.255000	5.630000	11.750000	12.820000	8.580000	0.510000	5.670000	18.860000	0.120000	0.600000	
50%	48.570000	95.680000	9.890000	21.690000	23.520000	15.850000	0.890000	9.160000	30.840000	1.070000	2.970000	
75%	80.590000	149.745000	19.950000	37.620000	40.127500	30.020000	1.450000	15.220000	45.570000	3.080000	9.150000	
max	949.990000	1000.000000	390.680000	362.210000	467.630000	352.890000	175.810000	193.860000	257.730000	455.030000	454.850000	1

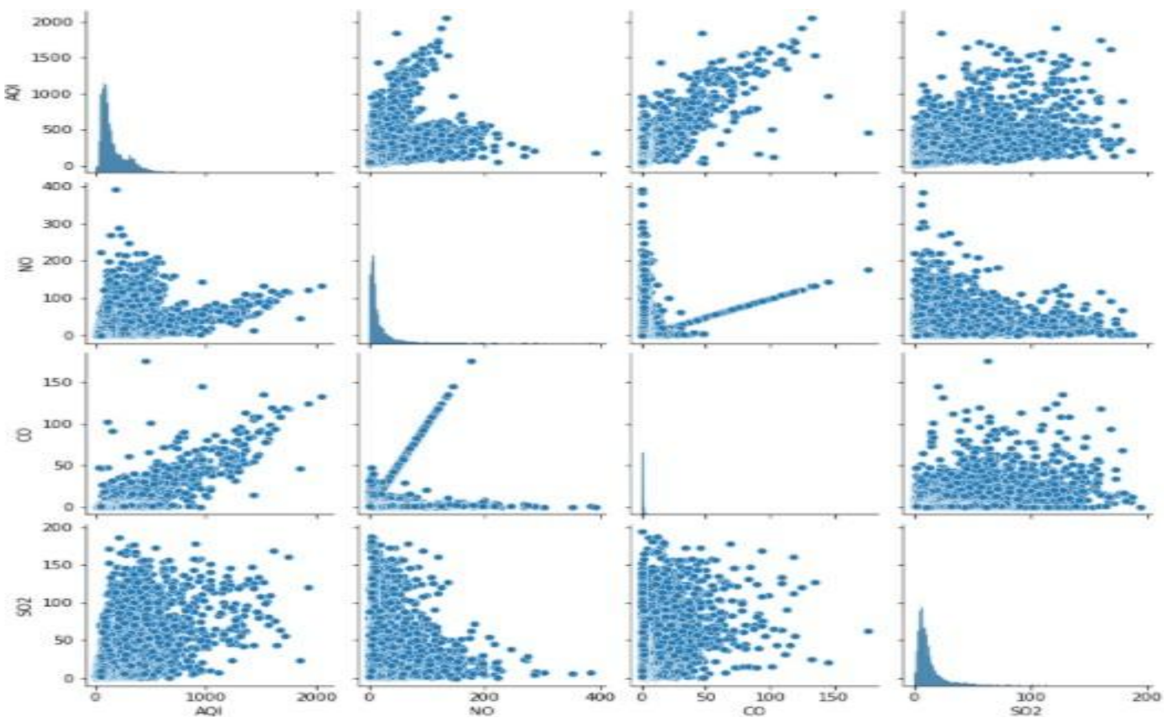
describe() method gives only the statistical document.

```
missing_val = pd.DataFrame(df.isna().sum()/len(df))
missing_val.columns = ['Proportion']
missing_val = missing_val.sort_values(by = 'Proportion', ascending=False)
missing_val.style.background_gradient(cmap='Blues')
```

	Proportion
Xylene	0.613220
PM10	0.377231
NH3	0.349734
Toluene	0.272290
Benzene	0.190410
AQI	0.158511
AQI_Bucket	0.158511
PM2.5	0.155701
NOx	0.141715
O3	0.136196
SO2	0.130507
NO2	0.121398
NO	0.121296
CO	0.069723

This data value shows amount of missing value of each constituent so for better result we either delete these constituents or use a mean value in place of null value.

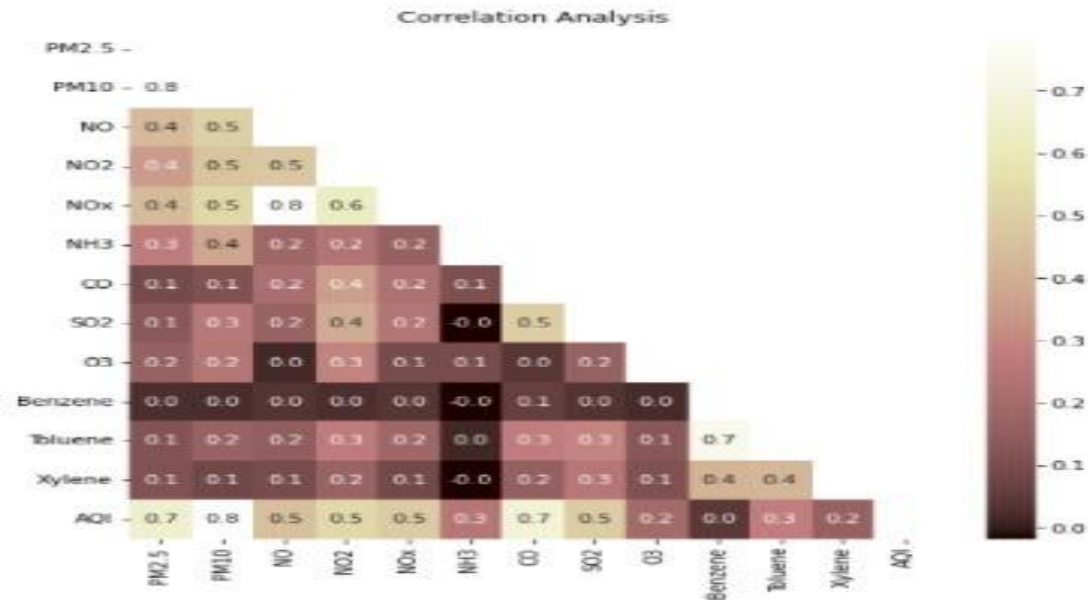
```
df2 = df[['AQI', 'NO', 'CO', 'SO2']]
sns.pairplot(df2)
```



AQI is for sure positively correlated with all the features in the plot i.e. SO2,NO,CO.

```
# FOr more details we can see for correlation heatmap of features

plt.figure(figsize=(8,8))
mask=np.triu(df.corr(method='pearson'))
sns.heatmap(df.corr(method='pearson'),
            annot=True,fmt='0.1f',
            mask=mask,
            robust=True,
            cmap='pink')
plt.title('Correlation Analysis')
```

```
Text(0.5, 1.0, 'Correlation Analysis')
```

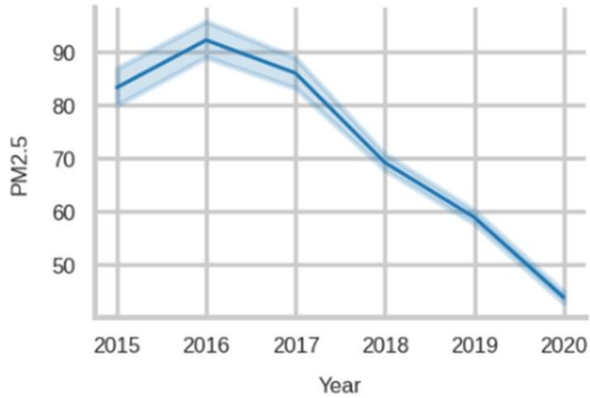


We transformed our data on monthly basis we lost our AQI_Bucket column in order to group all the numerical columns(features). So, we have included it back in our data.

It has a simple function of classifying the AQI into different buckets like normal, moderate, unhealthy, a hazardous, etc.

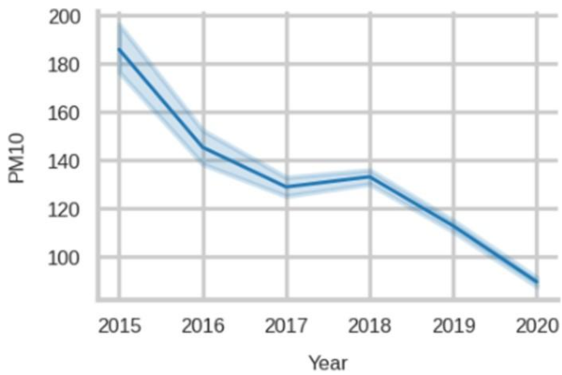
```
data2 = dataf.copy(deep=True)

conditions = [
               (data2['AQI'] > 0) & (data2['AQI'] <50),
               (data2['AQI'] > 51) & (data2['AQI'] <100),
               (data2['AQI'] > 101) & (data2['AQI'] <150),
               (data2['AQI'] > 151) & (data2['AQI'] <200),
               (data2['AQI'] > 201) & (data2['AQI'] <300),
               (data2['AQI'] > 300)
]

values = ['Good', 'Moderate', 'Unhealthy for sensitive group',
          'Unhealthy', 'Very Unhealthy', 'Hazardous']

data2['AQI_Bucket'] = np.select(conditions, values)

data2.head()
```

Some More Time-Analysis



PM2.5 year-wise

We can see the fall in the values PM2.5 as time increases on the x-axis.



PM10 year-wise

Here also the trends show a gradual depletion in levels.

Facts about our features

Here few things are common in all three plots:

All the values lie between 0-200 except PM2.5

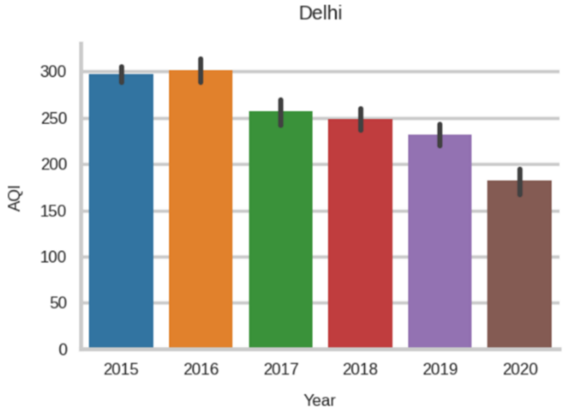
Range of counts for values < 100 varies in 10000s which is a healthy sign i.e mode of the distribution is centered around 50s in each.

Here we can see Nan values which will be handled later.

```
# url = "https://www.kaggle.com/rohanrao/air-quality-data-in-india?select=city_day.csv"
data = pd.read_csv('sample_data/city_day.csv')
df = pd.read_csv('sample_data/city_day.csv')
df.head()
```

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	AQI	AQI_Bucket
0	Ahmedabad	2015-01-01	NaN	NaN	0.92	18.22	17.15	NaN	0.92	27.64	133.36	0.00	0.02	0.00	NaN	NaN
1	Ahmedabad	2015-01-02	NaN	NaN	0.97	15.69	16.46	NaN	0.97	24.55	34.06	3.68	5.50	3.77	NaN	NaN
2	Ahmedabad	2015-01-03	NaN	NaN	17.40	19.30	29.70	NaN	17.40	29.07	30.70	6.80	16.40	2.25	NaN	NaN
3	Ahmedabad	2015-01-04	NaN	NaN	1.70	18.48	17.97	NaN	1.70	18.59	36.08	4.43	10.14	1.00	NaN	NaN
4	Ahmedabad	2015-01-05	NaN	NaN	22.10	21.42	37.76	NaN	22.10	39.33	39.31	7.01	18.89	2.78	NaN	NaN

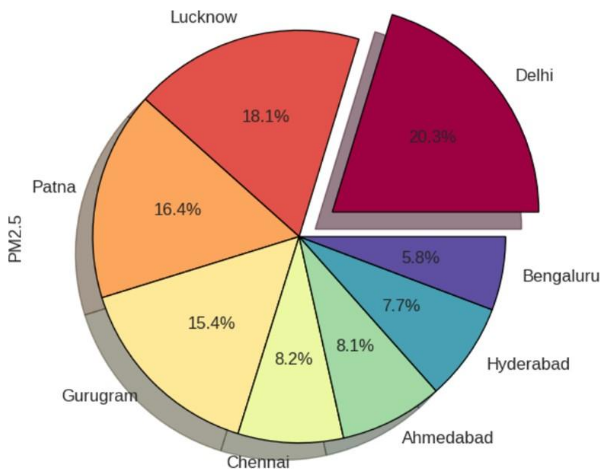
Some plots of analysis:
Delhi AQI Over Years



Points to be mentioned here are:

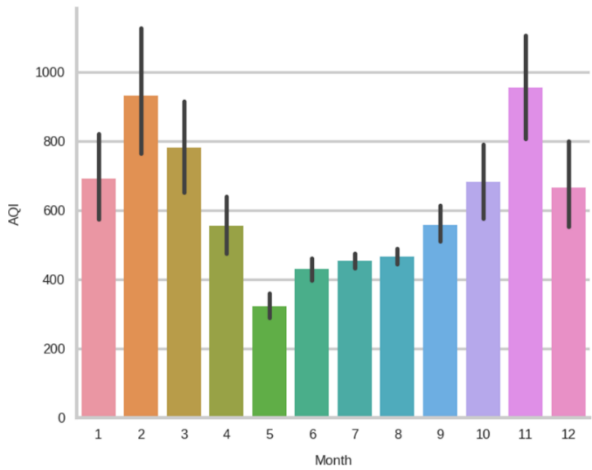
- Delhi is making progress over years as we can see by decreasing levels of AQI

Though the percentage progress is nice as per the plots, still the levels are in slight severity(i.e. above 175).

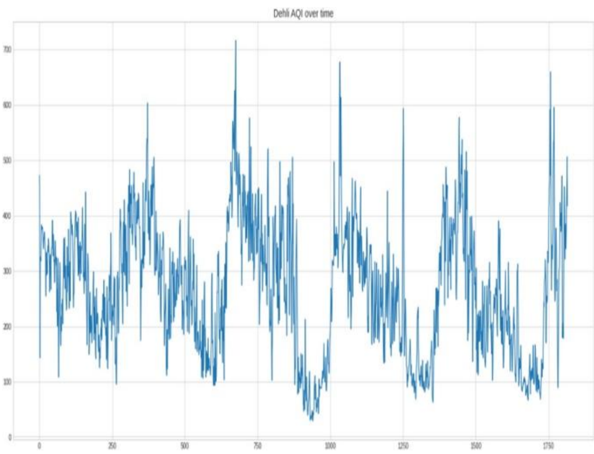


Here we can note some points à
•Delhi shares the biggest pie making it most polluted
•All the cities are capital regions which infers the respective other factors(like lifestyle, population, etc.) showing this trends.

MONTHLY AQI VALUES:

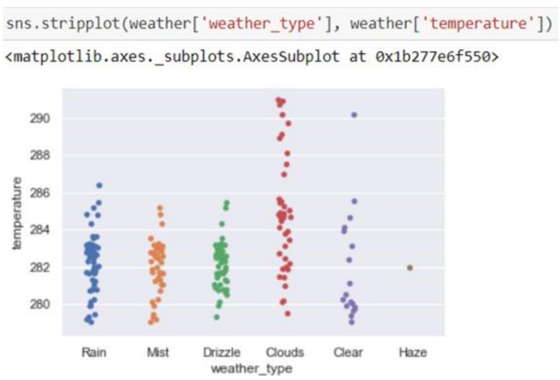


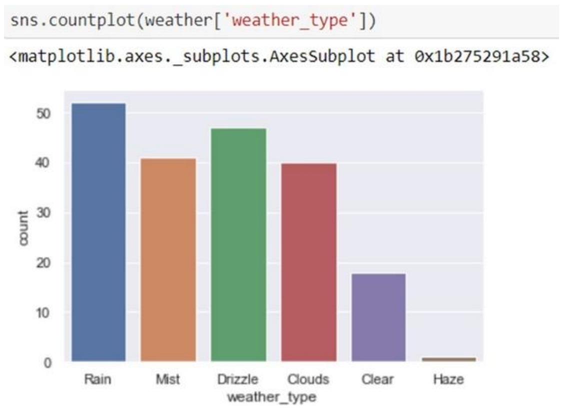
DELHI AQI VALUES:
X-AXIS:NO OF DAYS
Y-AXIS: AQI VALUE



PLOTS OF WEATHER CONDITIONS

Air pollution is largely affected by weather conditions like, windy conditions on a particular day can affect air quality in a much rapid way as compared to rainy day.





Some of the insights from our analysis are:

AQI is mostly affected by few of the features like pm2.5, pm10, co, no, so2.

Mostly capital cities have worse air quality as compared to small towns.

When AQI is high, temperature also rises.

Overs past 5 years average AQI value for some cities has decreased a bit, which is a healthy sign.

Still for most of the cities, average AQI values rise every year in a particular season.

Weather also plays an important role in defining air quality at any place.

IV. CONCLUSION

The Air Quality Index can provide a clear picture of the ambient air and the essential pollutants that are primarily responsible for air quality. The CPCB break point concentration was used to determine the AQIs. Particulate matter was shown to be the main cause of maximum times in the residential site NEERI, Nagpur, according to the AQI study. Particulate Matter is a severe global pollutant that is causing a decline in overall air quality. Particulates can come from a thermal power plant, a small-medium-scale company, or a vehicle, among other things. For the benefit of civic life, we must seek for proper pollution control and management plans such as planting and green belts. The use of this tool in development decision-making may be risky because it does not clearly address temporal AAQ variation due to meteorology, land use, ecosystem geology of the region and its impact, population exposure (poor) who cannot afford air conditioning comfort, chemical conversion and synergistic effect particle/gas combination leading to smoke acid rain and other climate change phenomena, health impact of raised AAQ due to agglomeration of smog, and population exposure (poor) who cannot afford air conditioning comfort.

REFERENCES

[1]. (PDF) Air Quality Index – A Comparative Study for Assessing the Status of Air Quality (researchgate.net)

[2]. Air Quality Index (AQI) Classification using CO and NO2 Pollutants: A Fuzzy-based Approach | IEEE Conference Publication | EEE Xplore

[3]. Sci-Hub | Research on Air Quality Forecasting Based on Big Data and Neural Network. 2020 International Conference on Computer Network, Electronic and Automation (ICCREA) | 10.1109/ICCNEA50255.2020.00045

[4]. https://www.kaggle.com/anvithashet/linearregeression

[5]. https://aqicn.org/data-platform/register/

[6]. https://towardsdatascience.com/real-time-dashboard-in-python-970e3c32b3d7

[7]. https://www.kaggle.com/vaishnavipatil4848/indian-air-quality-analysis

[8]. Air_pollution_Analysis/Test.csv at main · anuanmol/Air_pollution_Analysis · GitHub