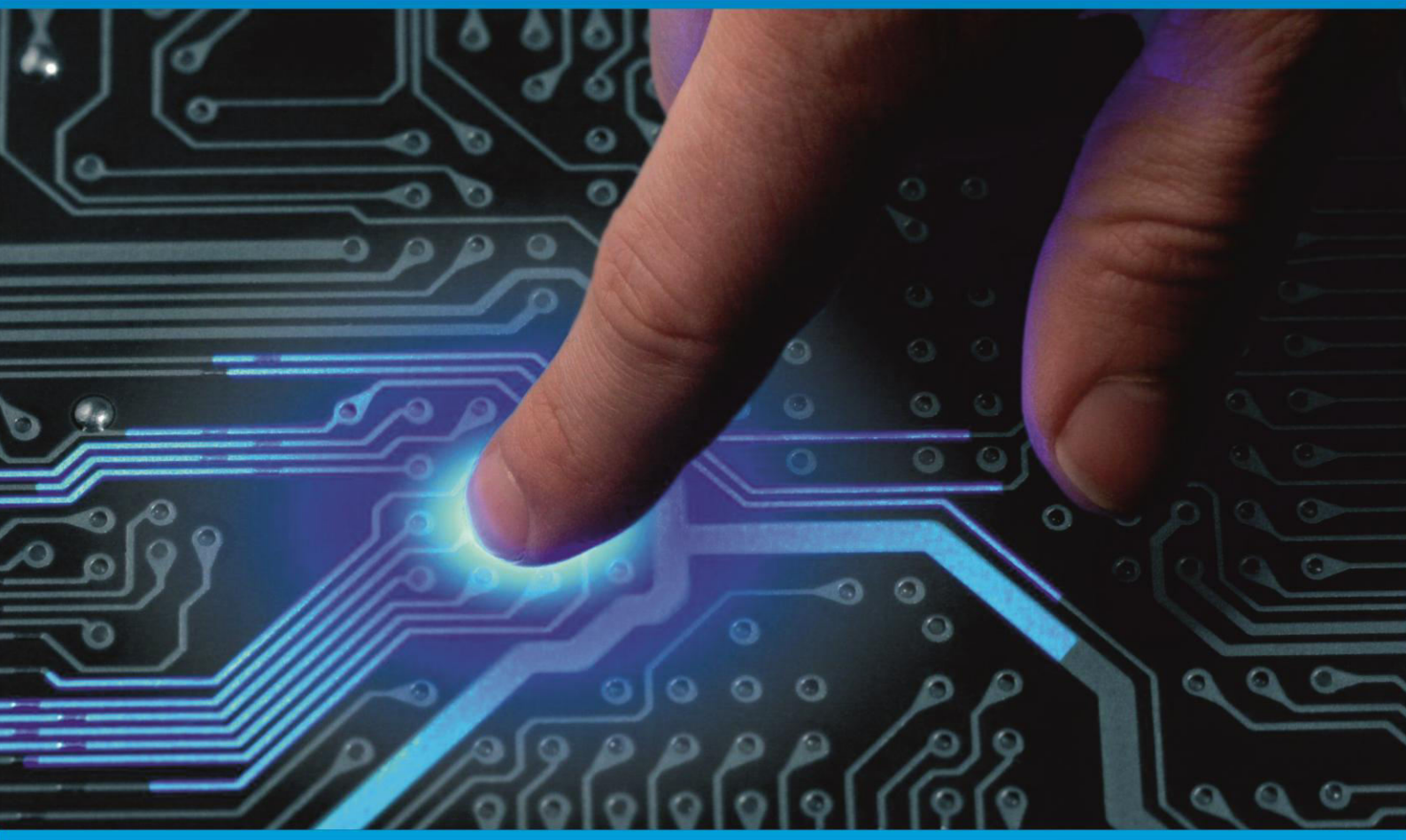
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905047 |

Statistical Analysis of India's Air Quality

Jasmeet Singh Ratra[1], Siddharth Nanda[2]

U.G Student, School of Engineering, Ajeenkya DY Patil University, Pune, Maharashtra, India [1]

Faculty, School of Engineering, Ajeenkya DY Patil University, Pune, Maharashtra, India [2]

ABSTRACT: Air pollution has been a major problem for very long in every nation and every move is used to tackle this problem. It is ranked as the sixth most dangerous killer in South Asia.One does not realize the harmful effects of a problem one has not experienced it in the first place. The following research paper introduces the reader about the air quality of different states in India to find some underlying principles or patterns which might give some insight into how severe the problem is. The paper will provide reader a base to see what aspects affected the quality of air, Example: Environment related Government Policy. The Expectation here is to study and analyse the changes and the reasons behind it.

KEYWORDS: Descriptive Statistics, Sampling, Dataset, Air quality, Analysis, NO_2, SO_2, Insights

I. INTRODUCTION

The Goal of the study is to analyse the air quality of different states of India from which different insights can be produced. The aim of the analysis will be the comparisons of each state of India- the country which ranges 3214km from north to south and 2,933 km from east to west. In South Asia, it is ranked as the sixth most dangerous killer.One does not realize the harmful effects of a problem one has not experienced it in the first place. India has seen major changes in its air quality since 90s. Python will be used for Descriptive statistics – helps to visualize data from a simple figure. In the following statistical analysis, samples from February 1990 to December 2015 will be visualized and following with the determination of the impacts made.

II. LITERATURE REVIEW

The research paper discuses about the range of quality of air in different regions that compares each state contributing to the air quality. Insights from 'The urban air quality' by Jesfringer has been derived, understood and studied for the following research.

Data Characteristics**:**For this study, the data taken from Historical Daily Ambient Air Quality Data is a cleaner version than the data released by the Ministry of Environment and Forests and Central Pollution Control Board of India under the National Data Sharing and Accessibility Policy (NDSAP).

The dataset contains the following features:

1. **stn_code:** Station code assigned to each station while recording data
2. **sampling_date**: Date of recording data
3. **state:**State name for measuring data
4. **location**: Represents the city whose air quality data is measured.
5. **agency:** Name of the agency that measured the data.
6. **type:** The type of area where the measurement was made.
7. **so2:** The amount of Sulphur Dioxide measured.
8. **NO2:** The amount of Nitrogen Dioxide measured
9. **rspm**: Respirable Suspended Particulate Matter measured.
10. **spm:** Suspended Particulate Matter measured.
11. **location_monitoring_station:** It indicates the location of the monitoring area.
12. **pm2_5:** It represents the value of particulate matter measured.
13. **date:** It represents the date of recording (It is a cleaner version of 'sampling_date' feature)

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905047 |

```
RangeIndex: 435742 entries, 0 to 435741
Data columns (total 13 columns):
 #   Column                       Non-Null Count   Dtype
---  ------                       --------------   -----
 0   stn_code                     291665 non-null  object
 1   sampling_date                435739 non-null  object
 2   state                        435742 non-null  object
 3   location                     435739 non-null  object
 4   agency                       286261 non-null  object
 5   type                         430349 non-null  object
 6   so2                          401096 non-null  float64
 7   NO2                          419509 non-null  float64
 8   rspm                         395520 non-null  float64
 9   spm                          198355 non-null  float64
10  location_monitoring_station  408251 non-null  object
11  pm2_5                        9314 non-null  float64
12  date                         435735 non-null  object
```

III. METHOD AND PROCEDURES

Features:

1)The 'type' feature: It represents the type of area where the data was recorded like industrial, residential etc. we will see the total number of such categories

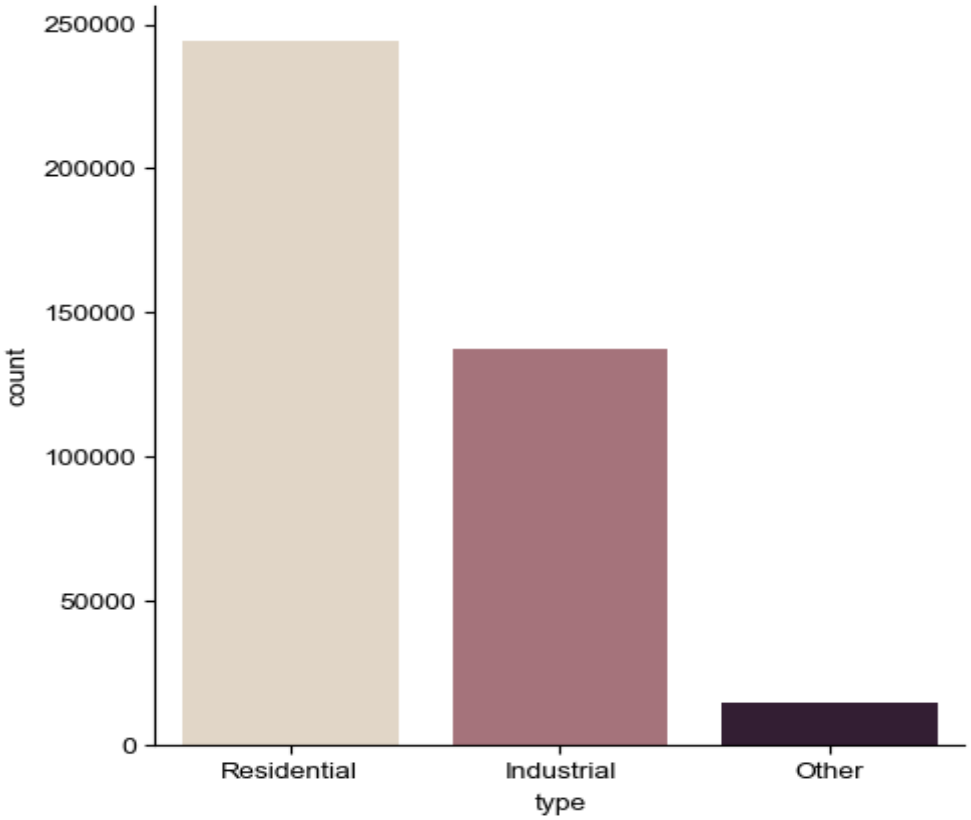
```
Residential, Rural and other Areas    179014
Industrial Area                        96091
Residential and others                 86791
Industrial Areas                       51747
Sensitive Area                          8980
Sensitive Areas                         5536
RIRUO                                   1304
Sensitive                                495
Industrial                               233
Residential                              158
```

For the following study only three types are considered i.e., Residential, Industrial and another feature. The following visualization will show the greatest number of a particular type of region.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905047 |



From the above figure one can say that the data was recorded with the focus near the residential area as it has the highest number of entries. This can be due to that the majority population lives near these 3 areas.

Sampling approach: Sampling is a first step to estimation. In the following dataset the samples used were taken from each month as per region of each state. In statistics there are two major approach to sample out the data – Simple random sampling and Stratified sampling. The following Data set uses Simple random sampling, and the reason is explained.

- In the case that any given region can only be selected once (i.e., after selection the data is removed from the selection pool):

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905047 |

$$P = 1 - \frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \cdots \cdot \frac{N-n}{N-(n-1)}$$

$$\overset{\text{Canceling:}}{=} 1 - \frac{N-n}{N}$$

$$= \frac{n}{N}$$

$$= \frac{100}{1000}$$

$$= 10\%$$

- In the case that any selected data is returned to the selection pool (i.e., can be picked more than once):

$$P = 1 - \left(1 - \frac{1}{N}\right)^n = 1 - \left(\frac{999}{1000}\right)^{100} = 0.0952\cdots \approx 9.5\%$$

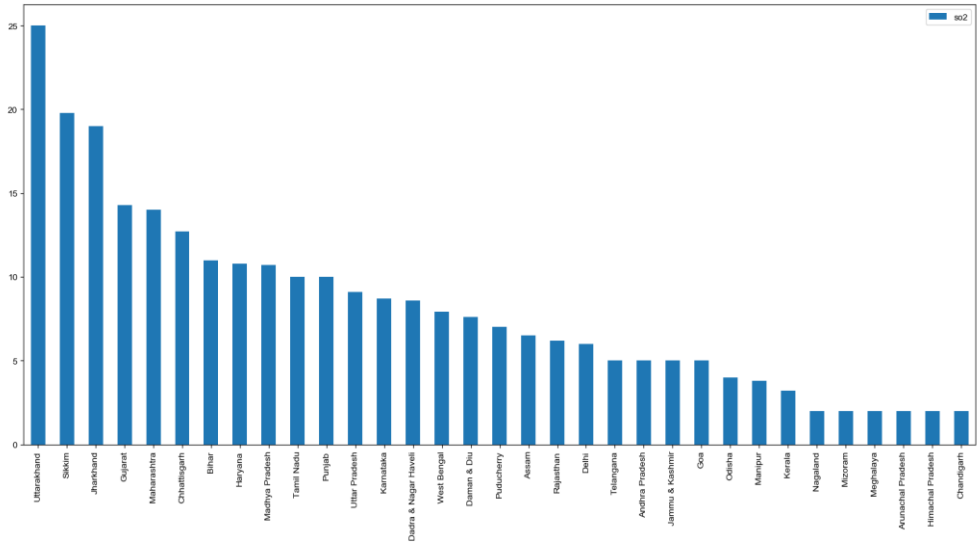
From a pool of huge recordings where the quality is a parameter and to find the relation between features would be ambiguous, thus a random sampling method fulfils the requirements.

Study:

After studying the dataset, we first compare each state by the amount of harmful pollutant.

On the x-axis will be the pollutant name. NO_2 and SO_2 levels will be studied in each state and regions.

Studying SO_2 concentration: SO_2 level from each state of India is compared using bar plot to see the visualize the differences easily.

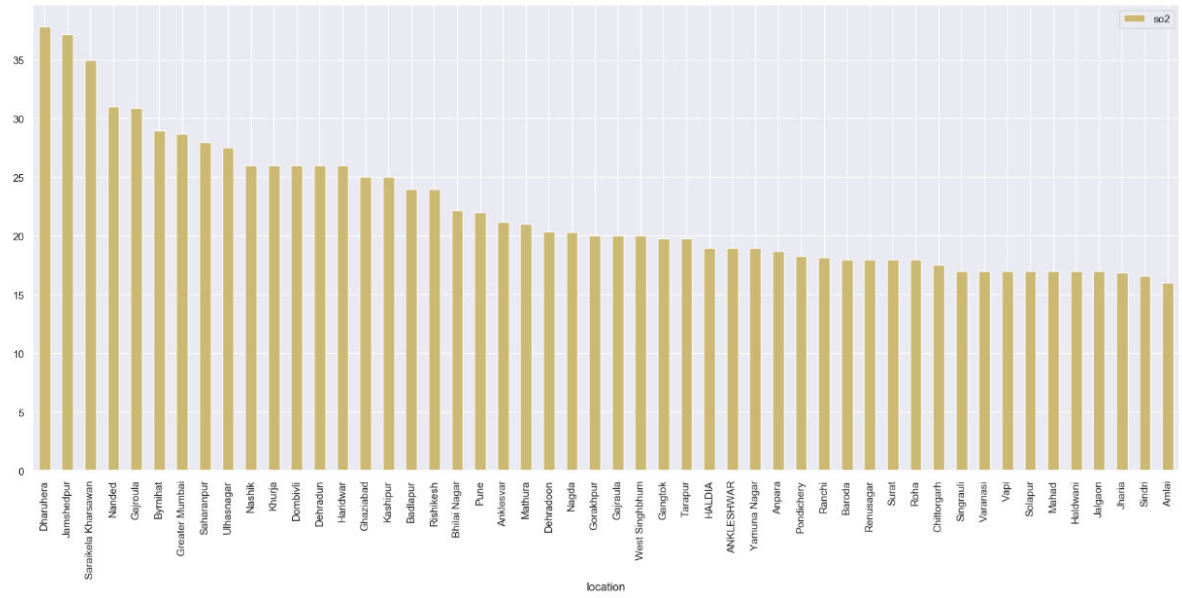


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

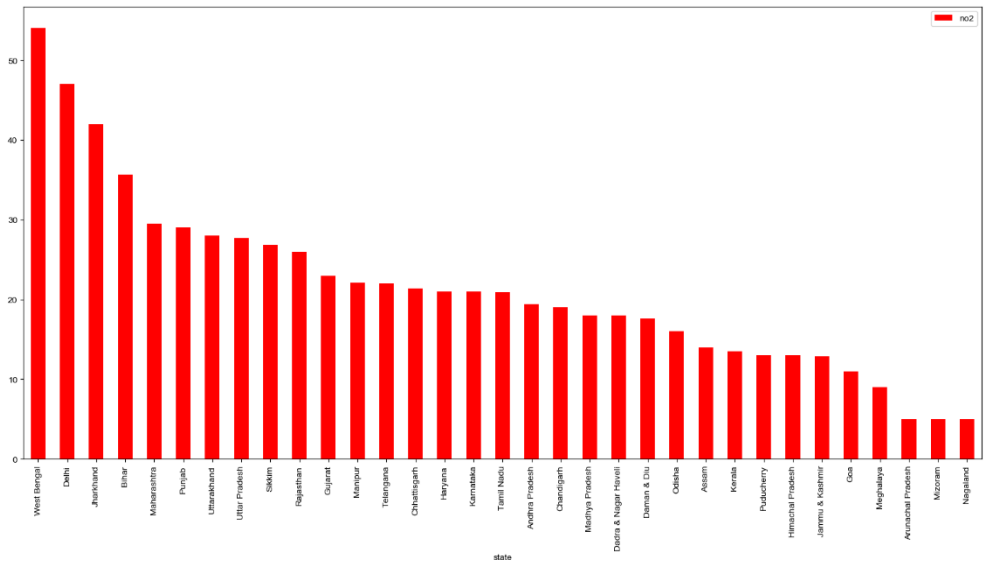
|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905047 |

Visualizing City wise per state: We will compare each city from each state to get a clearer idea which region affects the most to a state



Studying NO_2 concentration: NO_2 level from each state of India is compared using bar plot to see the visualize the differences easily.

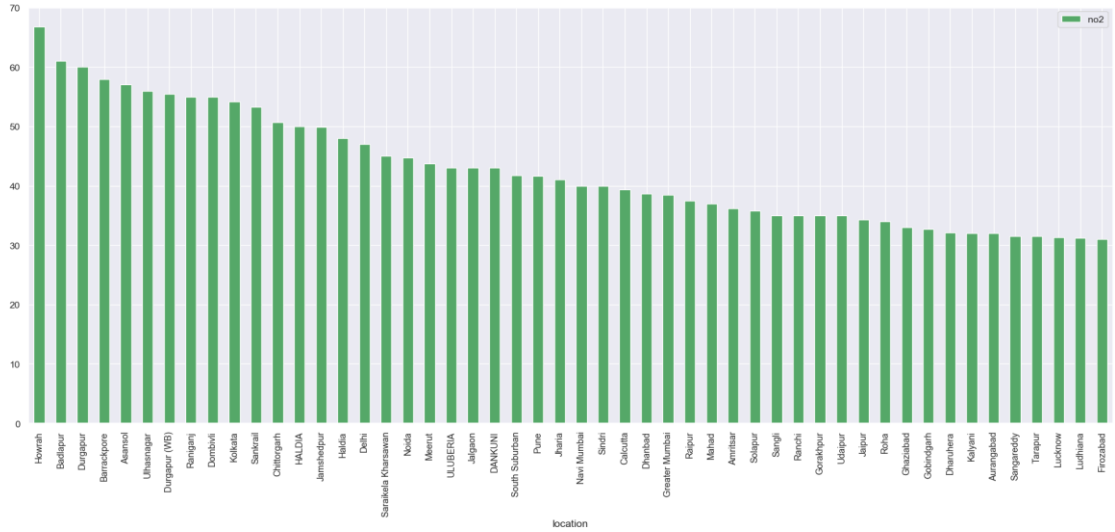


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

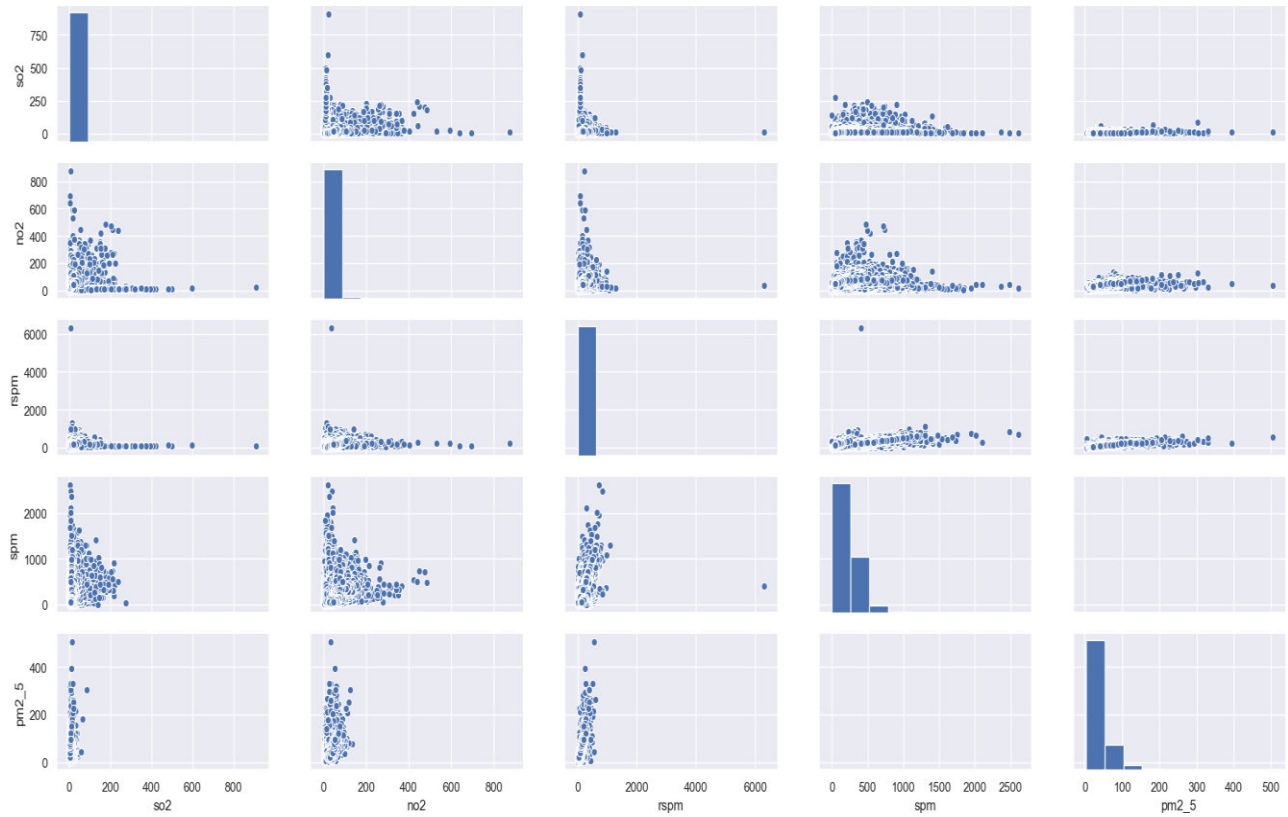
| DOI: 10.15680/IJIRCCE.2021.0905047 |

Visualizing City wise per state: We will compare each city from each state to get a clearer idea which region affects the most to a state



Statistical Analysis:

The paper only focuses on 2 major pollutants that is SO_2 and NO_2 because these two has the most adverse effects on the environment, but there are different pollutants which can be seen from the scatterplot



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905047 |

- For In-depth analysis, a corelation matrix will help to distinguish further

Summary of Visualizations:

The levels of SO_2 and NO_2 concentration were studied and compared in each state and the particular region. From the visualization it became clear that SO_2 level is highest in Uttarakhand and lowest in Chandigarh. We can Infer from this that states like Uttarakhand, Sikkim, Jharkhand, Gujarat, Maharashtra, Chhattisgarh have heavy levels of SO_2, and the reasons behind can be different for each state. When comparing region wise Dharuhera has the highest SO_2 concentration and is located in Haryana, followed by Jamshedpur, which is situated in Jharkhand, Madhya Pradeshand Sandra (Jharkhand), on the other hand, have the least concentrations of $SO_2$when compared to the other 50 locations. West Bengal leads the emissions of NO_2whereas Nagaland has the least amount of NO_2 present in the air. Delhi the capital of India comes second in the highest NO_2 concentration. The regions affected by mostNO_2 concentration is Raichur.

From the correlational matrix or heatmap, one can conclude that some states were heavily polluted in the early stages i.e. from 1980 to 2000 ,but later, there were visual changes in graphs which means some measures were taken.The reason for the decrease could be awareness in citizens and government policies. For example, The **Air (Prevention and Control of Pollution) Act, 1981** which was about prevention**,** control, and abatement of air pollution.

IV. FUTURE OF THE ANALYSIS

From the following performed analysis one can compare the air quality ranging from the late 90's to December 2015. This will help various groups of people; example government will see the effects of certain environment-basedregulations.Comparison will bring different perspective if certain objects are considered like Delhi reported a reduce in pollution after metro was built, thus metro was the object here and decrease in pollution was the result. So, such analysis will be important to take environment driven decisions in future.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905047 |

V. CONCLUSION

The Data analysis approachconcluded that data analysis is a crucial aspect for a better decisions and ideas. The approach towards the study was purely data-driven, however, was backed byvarious real-life instances. We saw that how data analysis and the day-to-day instances are coherent and how data analysis can be used to deal with significant problems like environment changes. The motive behind each analysis was to compare the air quality of India and what necessary steps shall be taken to improve it and gain various insights.

REFERENCES

1. Urban Air Quality.*JesFringer.*National Environmental Research Institute, Department of Atmospheric Environment, Frederiksborgvej 399, DK-4000 Roskilde.
2. Urban air quality.*Philip K. Hopke, David D. Cohen.* Science of The Total Environment, Volume 409,

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details