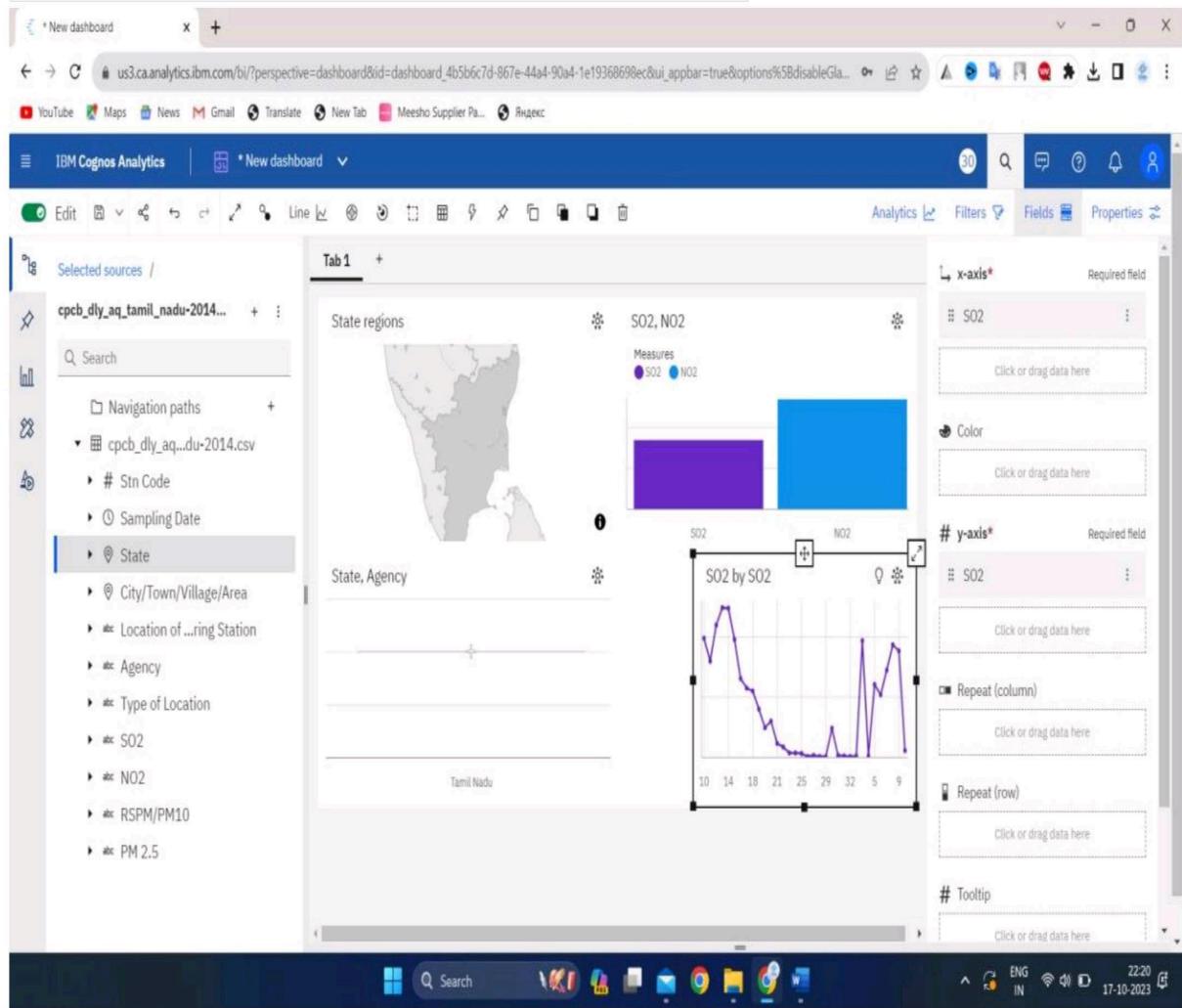


# Air Quality Analysis In TAMILNADU

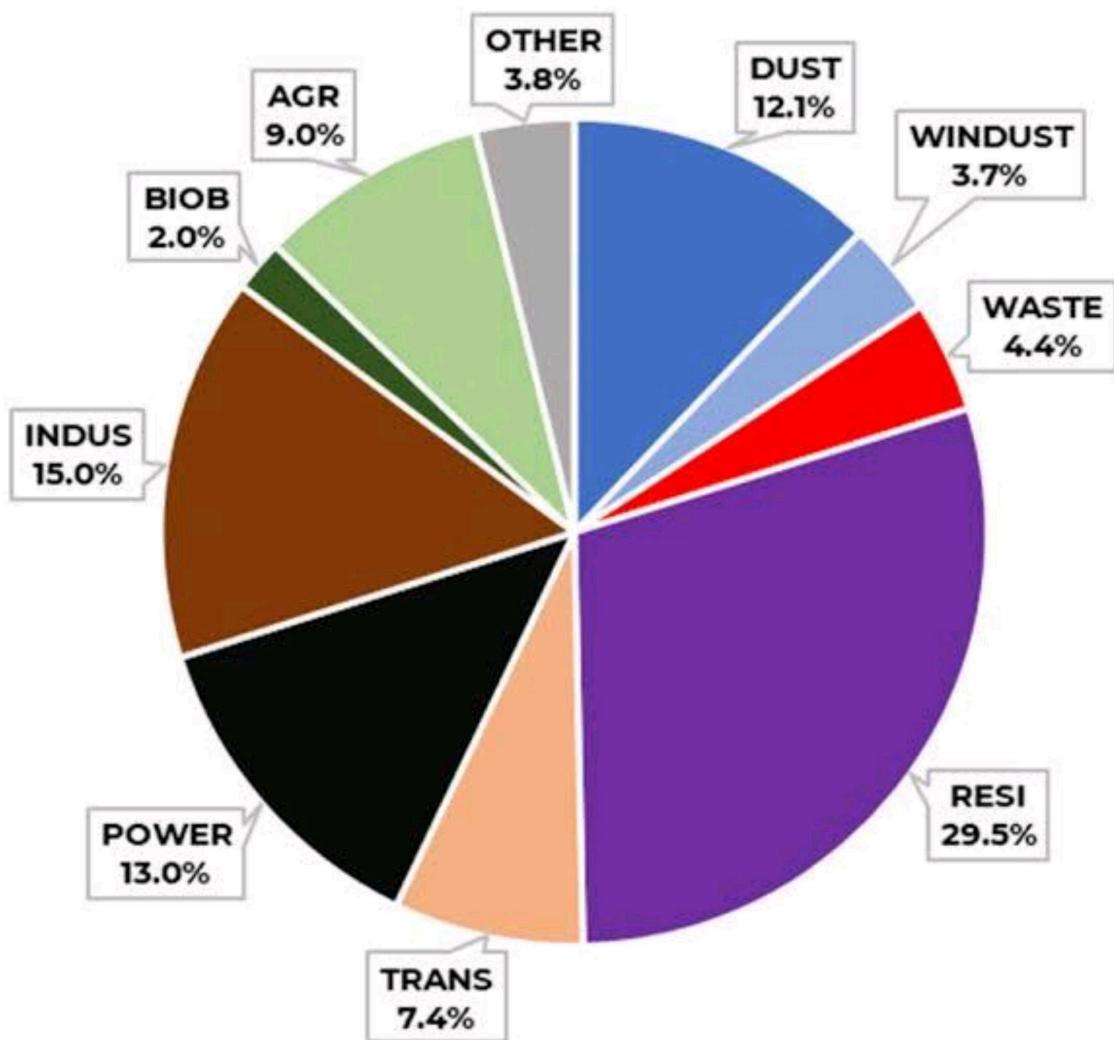
## With COGNOS

**Project Title: Customer Churn Prediction**

### IBM Cogon's Visualization :



## AIR Quality Analysis In Tamilnadu -Phase-3



## **INTRODUCTION**

- Clean air is the foremost requirement to sustain healthy lives of humankind and those of the supporting ecosystems which in return affect the human wellbeing. Due to the rapid growth of economy and fossil fuel and lack of emission controls, we have experienced substantially elevated.
- In an era where urbanization and industrialization are on the rise, monitoring and improving air quality has become a paramount concern for environmentalists, policymakers, and the general public. Poor air quality can have severe health, environmental, and economic consequences. To address this critical issue, we embark on an Air Quality Analysis Project utilizing the power of Python and various data manipulation libraries.
- This project aims to leverage the capabilities of Python, a versatile and widely-used programming language, along with popular libraries like NumPy, Pandas, Matplotlib, and Seaborn, to gather, analyze, and visualize air quality data. By harnessing the potential of these tools, we can better understand air quality patterns, identify sources of pollution, and contribute to data-driven decision-making for improving air quality standards.

## Given Dataset:

| STN CODE | SAMPLING DATE | STATE     | CITY    | MONITORING STATION | AGENCY                                  | TYPE OF LOCATION                   | SO2  | NO2  | RSPM/PM10 | PM2.5 |
|----------|---------------|-----------|---------|--------------------|---|------------------------------------|------|------|-----------|-------|
| 38       | 02-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 11   | 17   | 55 NA     |       |
| 38       | 07-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 13   | 17   | 45 NA     |       |
| 38       | 21-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 12   | 18   | 50 NA     |       |
| 38       | 23-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 15   | 16   | 46 NA     |       |
| 38       | 28-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 13   | 14   | 42 NA     |       |
| ....     | ....          | ....      | ....    | ....               | ....                                    | ....                               | .... | .... | ....      | ....  |
| 773      | 03-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 15   | 18   | 102 NA    |       |
| 773      | 10-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 12   | 14   | 91 NA     |       |
| 773      | 17-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 19   | 22   | 100 NA    |       |
| 773      | 24-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 15   | 17   | 95 NA     |       |
| 773      | 31-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 14   | 16   | 94 NA     |       |

The above is the given dataset of Air Quality Analysis in Tamil Nadu in 2014. This dataset consists of 11 columns and 2880 rows.

## NECESSARY STEPS TO FOLLOW

### 1.Import Libraries:

Start by importing the neccessary libraries. Load your dataset into the pandas dataframe. And then display the output.

## Program:

```
# importing pandas module for data frame
import pandas as pd

# loading dataset and storing in train variable
train=pd.read_csv('cpcb_dly_aq_tamil_nadu-2014.csv')

# display top 5 data
train.head()
```

## Output:

| STN CODE | SAMPLING DATE | STATE     | CITY    | MONITORING STATION | AGENCY                                  | TYPE OF LOCATION | SO2 | NO2 | RSPM/PM10 | PM2.5 |
|----------|---------------|-----------|---------|--------------------|---|------------------|-----|-----|-----------|-------|
| 38       | 02-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA  | 11  | 17  | 55        | NA    |
| 38       | 07-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA  | 13  | 17  | 45        | NA    |
| 38       | 21-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA  | 12  | 18  | 50        | NA    |
| 38       | 23-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA  | 15  | 16  | 46        | NA    |
| 38       | 28-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA  | 13  | 14  | 42        | NA    |

## 2.Exploratory Data Analysis (EDA):

Perform EDA to understand your data better. This includes checking for missing values, exploring the data's statistics, and visualizing it to identify patterns.

## **PROGRAM:**

```
import pandas as pd import
numpy as np import seaborn
as sns import matplotlib.pyplot
as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score,
mean_absolute_error, mean_squared_error from
sklearn.linear_model import LinearRegression from
sklearn.linear_model import Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR import xgboost as xg
%matplotlib inline import warnings
warnings.filterwarnings("ignore")
/opt/conda/lib/python3.10/site-packages/scipy/_init_.py:146:
UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required
for this version of SciPy (detected version 1.23.5
warnings.warn(f'A   NumPy   version   >={np_minversion}   and
<{np_maxversion} "
```

## Loading dataset:

```
dataset = pd.read_csv('cpcb_dly_aq_tamil_nadu-2014.csv')
```

## OUTPUT:

| STN CODE | SAMPLING DATE | STATE     | CITY    | MONITORING STATION | AGENCY                                  | TYPE OF LOCATION                   | SO2  | NO2  | RSPM/PM10 | PM2.5 |
|----------|---------------|-----------|---------|--------------------|---|------------------------------------|------|------|-----------|-------|
| 38       | 02-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 11   | 17   | 55        | NA    |
| 38       | 07-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 13   | 17   | 45        | NA    |
| 38       | 21-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 12   | 18   | 50        | NA    |
| 38       | 23-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 15   | 16   | 46        | NA    |
| 38       | 28-01-2014    | TAMILNADU | CHENNAI | KATHIVAKKAM        | TAMILNADU STATE POLLUTION CONTROL BOARD | INDUSTRIAL AREA                    | 13   | 14   | 42        | NA    |
| ....     | ....          | ....      | ....    | ....               | ....                                    | ....                               | .... | .... | ....      | ....  |
| 773      | 03-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 15   | 18   | 102       | NA    |
| 773      | 10-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 12   | 14   | 91        | NA    |
| 773      | 17-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 19   | 22   | 100       | NA    |
| 773      | 24-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 15   | 17   | 95        | NA    |
| 773      | 31-12-2014    | TAMILNADU | TRICHY  | CENTRAL BUS STAND  | TAMILNADU STATE POLLUTION CONTROL BOARD | RESIDENTIAL ,RURAL AND OTHER AREAS | 14   | 16   | 94        | NA    |

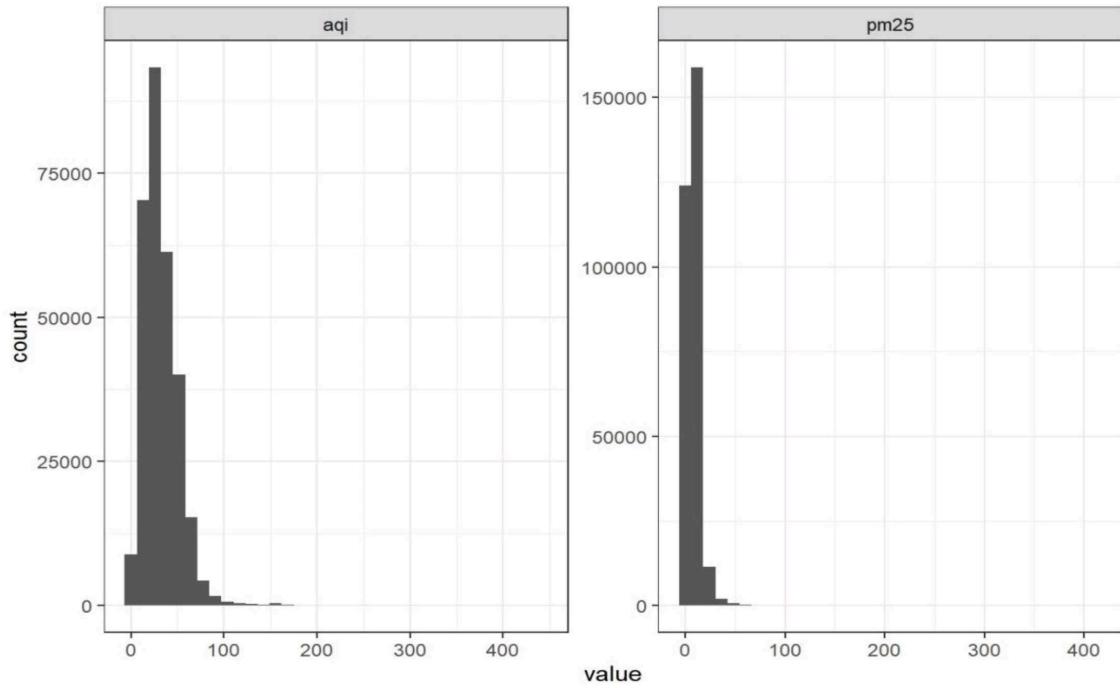
## PREPROCESSING THE DATASET

- Data preprocessing is the process of cleaning, transforming, and integrating data in order to make it ready for analysis.
- Preprocessing often involves feature selection, creation, and transformation. This crucial step allows us to identify the most relevant variables, reduce dimensionality, and adapt the data to better suit our modeling goals.

## Program 1:

```
## histogram for aqi and pm2.5  
vis <-  
  pmaqi %>%  
    select(pm25, aqi) %>%  
    gather(key = class, value = value)  
  ggplot(data = vis) +  
    geom_histogram(aes(x = value), bins = 35) +  
    facet_wrap(~ class, scales = "free") +  
    theme_bw()
```

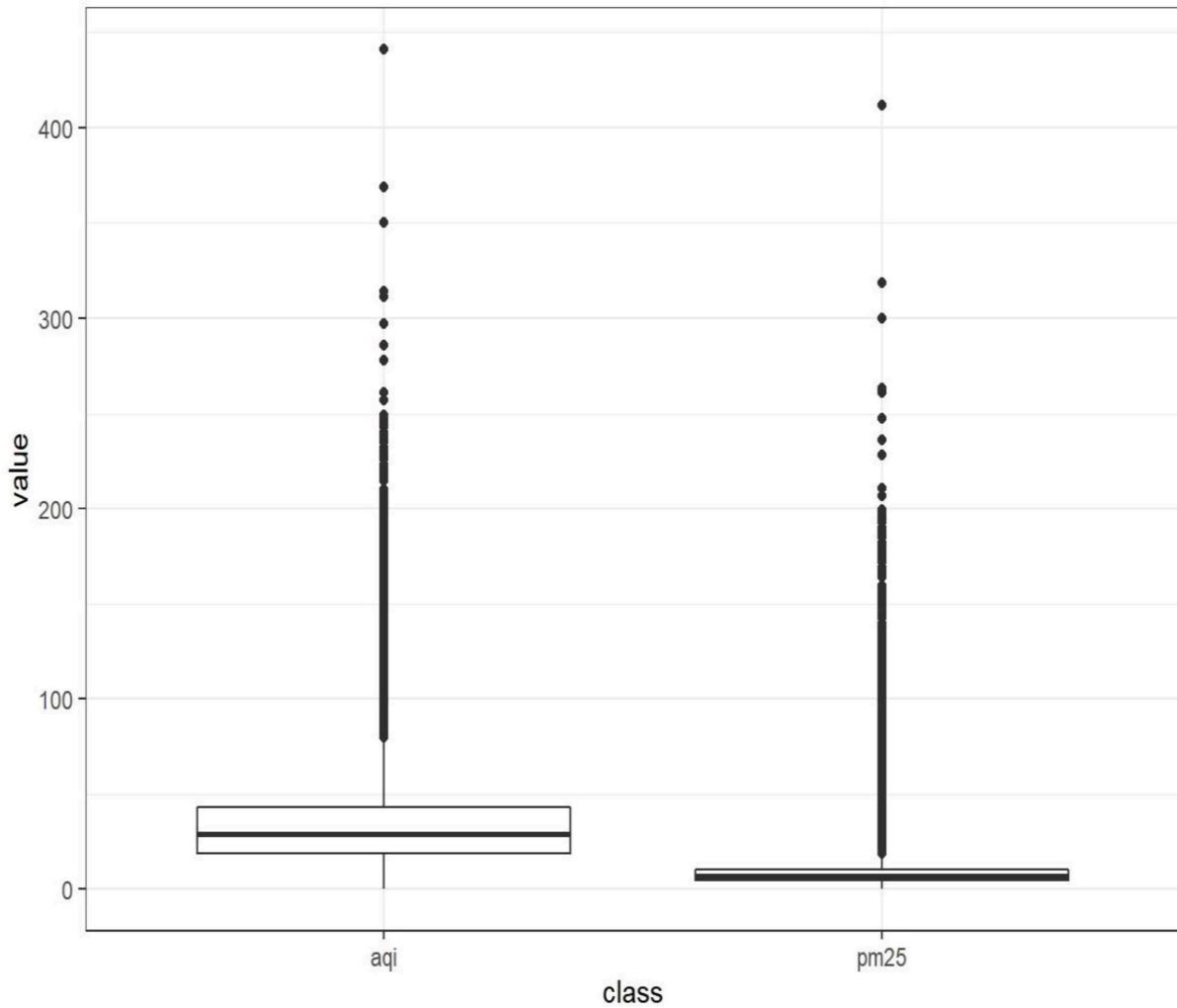
## Output:



## Program 2:

```
ggplot(data = vis) +  
  geom_boxplot(aes(x = class, y = value)) +  
  theme_bw()
```

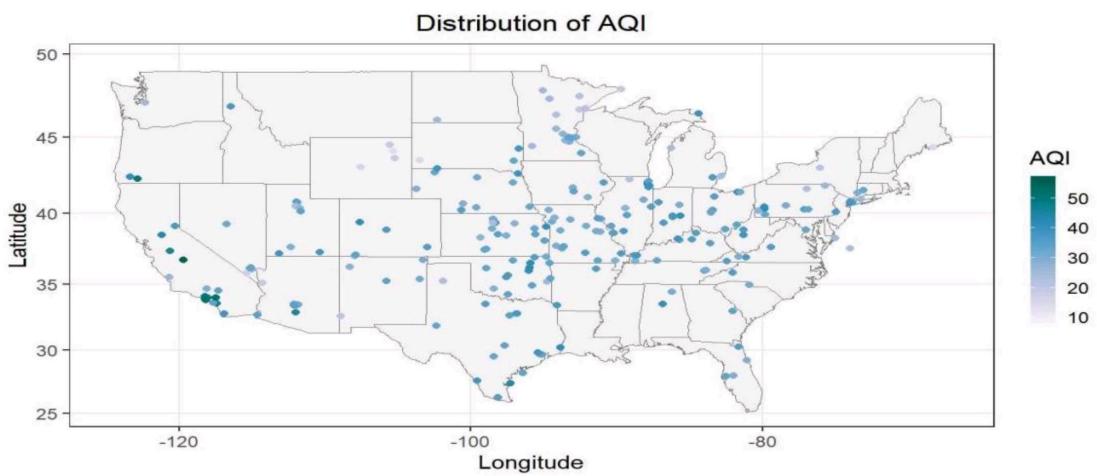
## Output:



## Program 3:

```
#for aqi index
vis2 <- vis[vis$class == "aqi", ]
ggplot() +
  geom_polygon(aes(x = long, y = lat, group = group),
               fill = "whitesmoke", colour = "gray50", size = 0.1,
               data = states) +
  geom_point(aes(x = long, y = lat, color = value), data = vis2) +
  scale_color_distiller("AQI", palette = "PuBuGn", direction = 0) +
  xlab("Longitude") + ylab("Latitude") +
  ggttitle("Distribution of AQI") +
  coord_map() +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        legend.key.width = unit(4, "mm"))
```

## Output:



# Challenges in Tamil Nadu

## **Urbanization Pressure:**

- Rapid urban growth leads to increased vehicular emissions, industrial pollution, and construction activities, all contributing to poor air quality.

## **Industrial Expansion:**

- Tamil Nadu's industrial sectors, including manufacturing and power plants, emit pollutants. Balancing industrial growth with environmental conservation is a challenge.

## **Agricultural Practices:**

- Agricultural activities, including crop burning, can contribute to particulate matter in the air, impacting rural and urban areas alike.

## **Population Density:**

- High population density in cities intensifies pollution. Residential areas, vehicular traffic, and commercial zones coexist closely, leading to concentrated emissions.

## **CONCLUSION**

- Loading and preprocessing the dataset are pivotal phases in any data analysis project, and they hold equal importance in the context of an air quality analysis project. These initial steps are instrumental in shaping the trajectory of the project and the quality of the insights that can be extracted.
- By loading the dataset, we establish the foundation for data analysis and machine learning. It allows us to confirm that the data is accessible and formatted correctly, laying the groundwork for reliable analysis. Preprocessing encompasses data cleaning, where missing values, duplicate records, and outliers are addressed. Clean data is the cornerstone of accurate analysis and modeling.
- In conclusion, dataset loading and preprocessing are not merely preliminary chores but the cornerstones of successful air quality analysis. They set the stage for rigorous, reliable, and informative analysis, offering the potential to yield actionable insights that can contribute to the betterment of air quality, public health, and environmental well-being. The care and precision applied to these phases directly influence the strength and significance of the project's outcomes.

