Air quality analysis in Tamilnadu

# Introduction:

Poor air quality has severe health implications, leading to respiratory diseases, cardiovascular issues, and other health problems. Air pollution not only affects human health but also harms the environment. It can lead to acid rain, damage to ecosystems, and the deterioration of historical monuments. Tamil Nadu has taken steps to monitor and control air pollution. Government agencies such as the Tamil Nadu Pollution Control Board (TNPCB) oversee air quality regulations and emission standards. Continuous monitoring stations are set up across the state to collect air quality data. Air quality analysis relies on advanced technologies, including air quality monitoring stations, satellite imagery, and data analytics. Innovations in sensor technology and data integration are helping to provide real-time air quality information.

# About phase 3:

In this part we will begin building your project by loading and preprocessing the dataset. Begin the analysis by loading and preprocessing the air quality dataset.

Load the dataset using Python and data manipulation libraries (e.g., pandas).

# About IBM cognos:

Cognos is a business intelligence performance management tools for IBM that allows technical and non-technical employees in any company to analyse, extract and create interactive dashboards that enable the company to take relevant key decisions. The Cognos is an intelligence-gathering platform for business that provides an analytical solution for business needs that is scalable and self-service. The highly interactive nature makes it a good way of creating user-friendly dashboards and reports for every company.

# Data loading and pre-processing:

To begin the analysis by loading and preprocessing an air quality dataset using Python and data manipulation libraries, you can follow these general steps:

1. Import Necessary Libraries: Start by importing the necessary Python libraries for data analysis and manipulation. Common libraries include Pandas for data handling and NumPy for numerical operations.

2. Load the Dataset: Use Pandas to load the air quality dataset. You can read data from various file formats, such as CSV, Excel, or databases. Replace `'your\_dataset.csv'` with the actual file path or URL of your dataset.

3. Explore the Dataset: Take an initial look at the dataset to understand its structure. You can use methods like `head()`, `info()`, and `describe()`.

4. Data Preprocessing: Perform data preprocessing steps, which may include handling missing values, data type conversions, and data cleaning. For example, if there are missing values, you can use Pandas methods like `fillna()` to fill them.

These steps provide a general framework for loading and preprocessing an air quality dataset in Python. The specific steps and libraries you use will depend on the dataset's structure and your analysis goals.

# Visualization using IBM cognos:

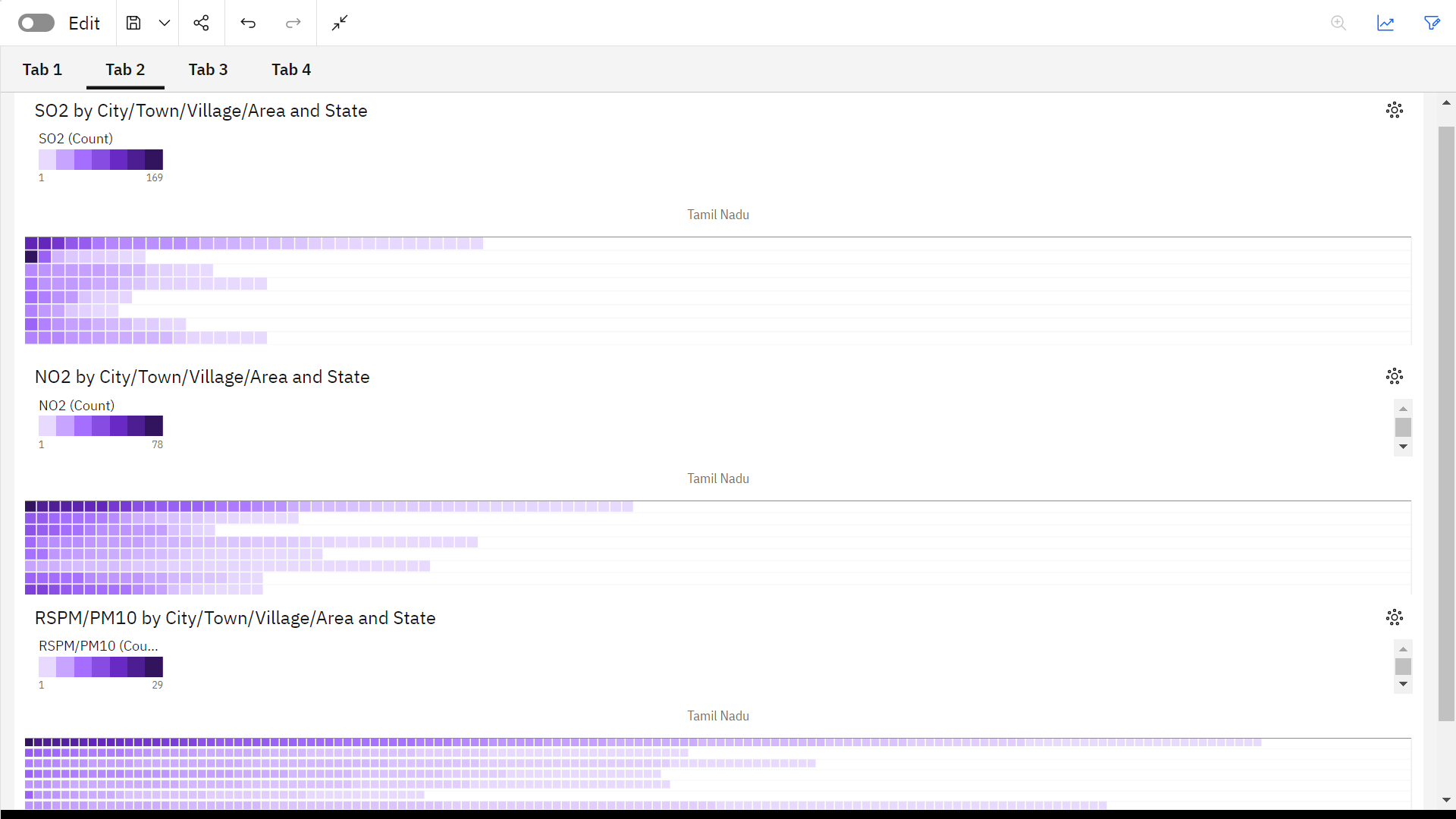
## Histogram:

A histogram is a graphical representation of the distribution of data. It's commonly used to visualize the frequency or count of data points within predefined intervals or "bins." The x-axis represents the data range or values, and the y-axis represents the frequency or count of data points falling into each bin.

# 

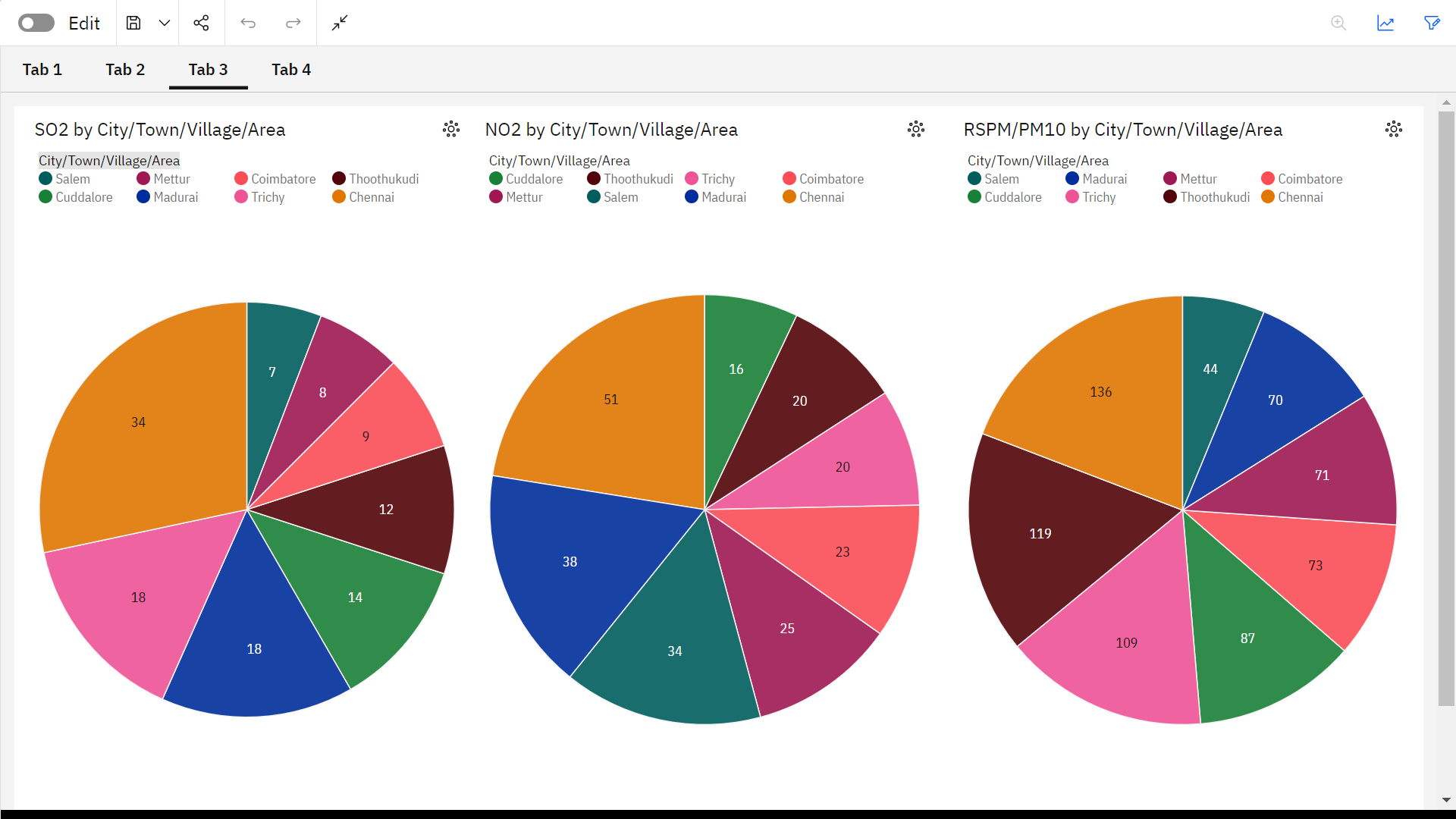
## Heatmap:

A heatmap is a graphical representation of data where individual values are represented as colors. It is a way to visualize data in a two-dimensional space, with each data point or value represented by a colored cell. Heatmaps are particularly useful for showing patterns, correlations, and variations in data, making them easier to understand and interpret.



## Pie chart:

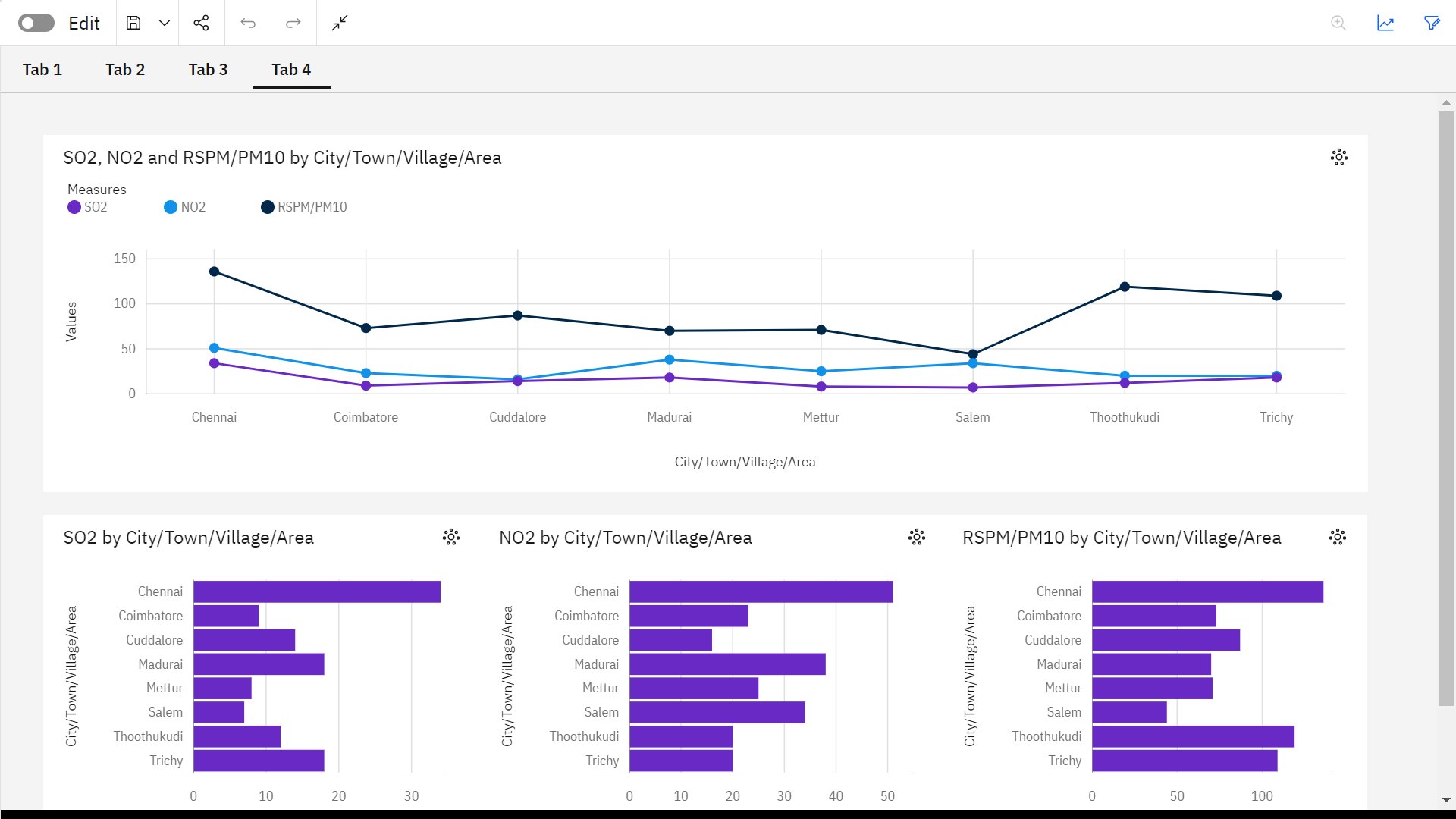
A pie chart is a circular statistical graphic that is used to represent data in a simple and visually appealing way. It is a type of chart that divides a circle into sectors or "slices," where each sector represents a proportion of the whole. The size of each sector is typically proportional to the quantity it represents. Pie charts are particularly useful for showing the composition or distribution of a data set.



## Line chart and bar chart:

A bar chart is a graphical representation of data using rectangular bars or columns of varying heights. Each bar typically represents a category or group, and the height of the bar is proportional to the value it represents. Bar charts can be either vertical (column chart) or horizontal (bar chart), depending on the orientation of the bars.

A line chart is a type of data visualization that is used to represent data points over a continuous interval or time period, and it connects these data points with straight lines. It is particularly useful for showing trends, changes, or fluctuations in data over time. Each data point in a line chart is usually represented as a marker (e.g., a point or circle) at specific coordinates, and the markers are connected by lines to form a continuous representation of the data.



## Conclusion:

The analysis of air quality data involves several important steps, including data loading, preprocessing, visualization, statistical analysis, and machine learning. The specific steps and libraries used may vary depending on the dataset and analysis goals. The findings from the analysis can provide valuable insights for decision-making and improving air quality.