**AIR QUALITY ASSESSMENT TAMIL NADU**

**INTRODUCTION:**

As part of the initiative, data from Tamil Nadu monitoring stations will be examined and shown. Understanding the amount of air pollution and developing a forecast model to determine RSPM/PM10 levels based on amounts of NO2 and SO2. The objectives of this project are established, and the analytical plan is a design is chosen, visualization techniques are used, and a prediction model is constructed.utilizing Python and the necessary libraries.

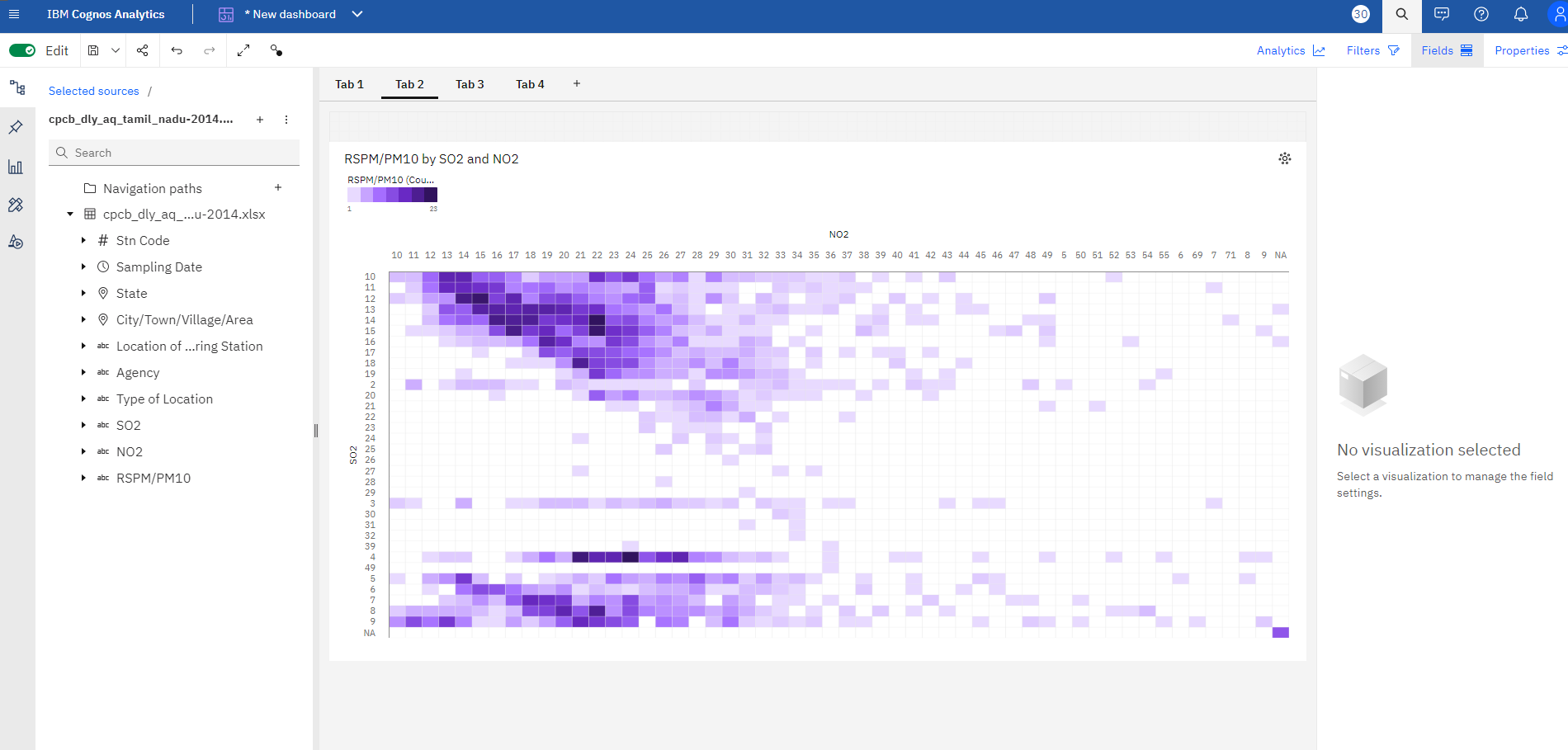
**DESCRIPTION:**

In this technological project, you will carry out various analyses, model constructions, and evaluations in accordance with the project specifications. Utilise IBM Cognos to carry out various analyses and visuals. After completing the pertinent tasks, write a document about it and distribute it for evaluation.

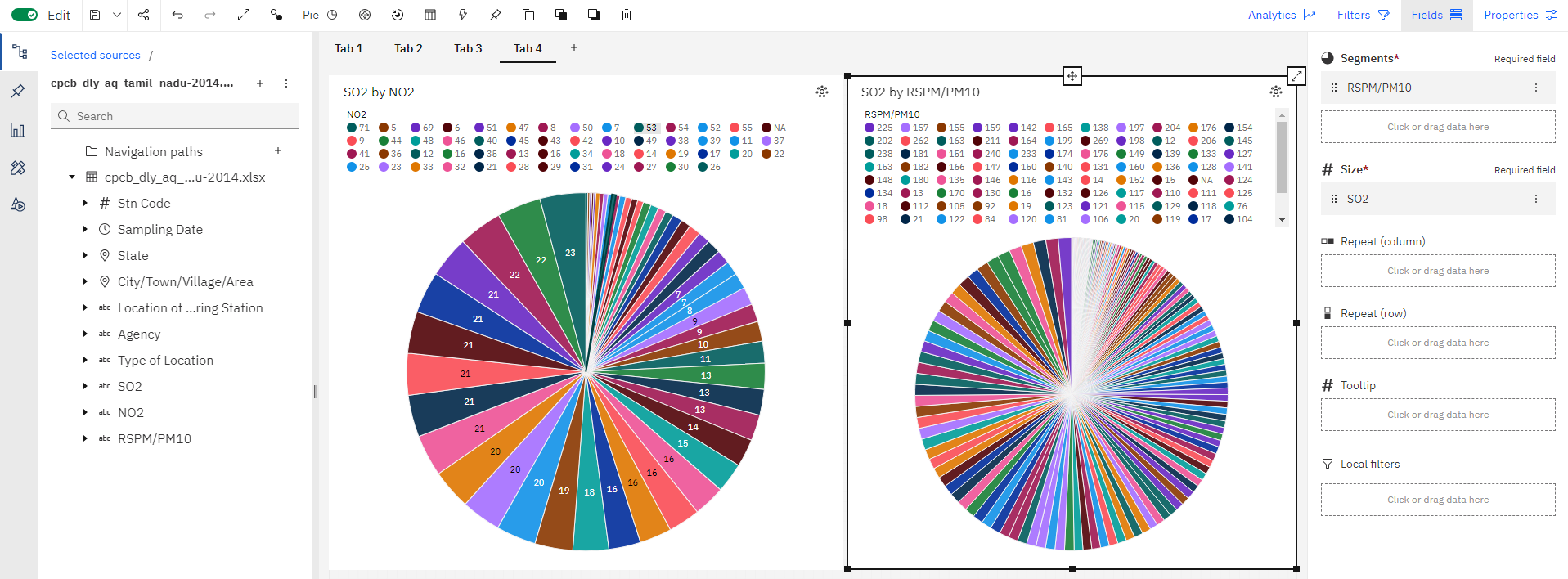
**IBM COGNOS TOOL:**

The IBM Cognos tool is used for analyzing the files such as csv files and other files to visualize data from them.

* **HEATMAP:**



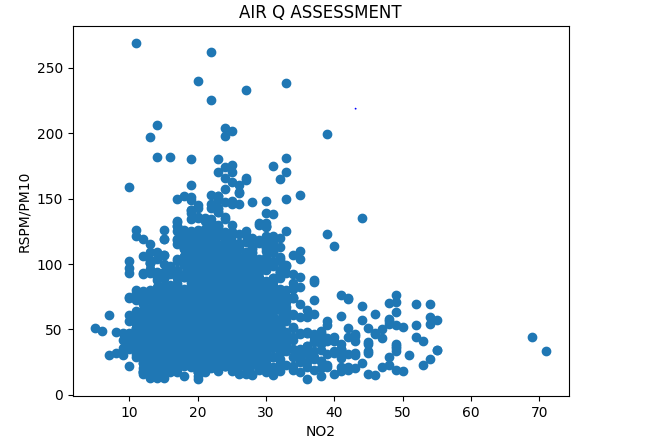
* 13 has a RSPM/PM10 of 20 for Stn Code 366.
* 4 has a RSPM/PM10 of 45 for Stn Code 375
* SO2 14 has the highest RSPM/PM10 at 215, out of which NO2 22 contributed the most at 23.
* **PIE CHART:**

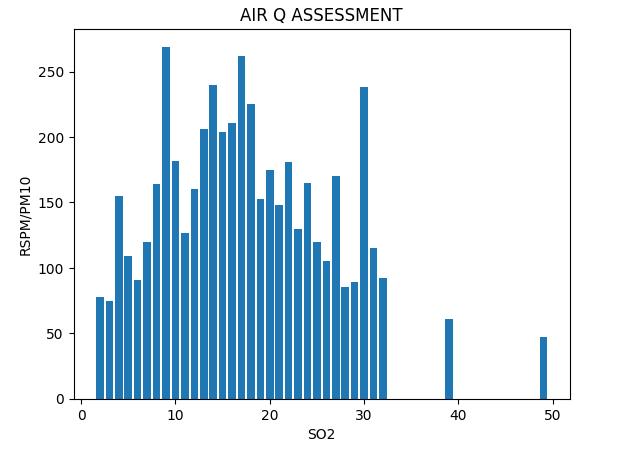


* 29 has a SO2 of 6 for Stn Code 161.
* Stn Code 767 has the highest SO2 at 108, out of which RSPM/PM10 95 contributed the most at 4.
* RSPM/PM10 41 SO2 from Stn Code 71 is 4, whereas 44 is only 1.
* 47 (2.2 %) and 41 (2.2 %) are the most frequently occurring categories of RSPM/PM10 with a combined count of 126 items with SO2 values (4.4 % of the total).

**VISUALIZATION USING PYTHON:**

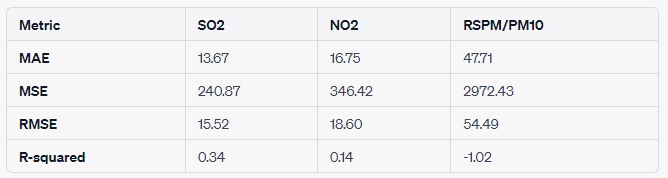
Matplotlib and Seaborn are two examples of libraries used in Python visualization. A multitude of charts and graphs, including bar plots, scatter plots, and heatmaps, can be made to present data in an efficient manner. Python is a popular choice for data visualization because of its versatility and large library. Data analysis and communication can be effectively facilitated by customizing visualizations, adding labels, and creating interactive plots for data presentation and exploration.





**PERFORMANCE METRICS:**

Accuracy is a statistic indicating the accuracy of measurements or predictions, calculated by dividing the total number of measurements by the number of accurate predictions, with low errors indicating high accuracy. Tamil Nadu assesses air quality using important performance criteria, such as CO, NO2, SO2, PM2.5, and ozone. By classifying contaminants, the Air Quality Index enables government agencies to take appropriate measures to safeguard public health and lower pollution levels.



Let's delve into the metrics and what they convey about the performance of a regression model using the provided data,

* **Mean Absolute Error (MAE):**
  + **SO2**: The average absolute difference between the predicted and actual SO2 values is approximately 13.67. This represents the typical magnitude of error in the model's predictions for SO2.
  + **NO2:** For NO2, the MAE is around 16.75, indicating a slightly larger average absolute error compared to SO2.
  + **RSPM/PM10:** The model's predictions for RSPM/PM10 have an average absolute error of about 47.71.
* **Mean Squared Error (MSE):**
  + **SO2:** The MSE for SO2 is 240.87, reflecting the average squared difference between predicted and actual SO2 values. A lower MSE indicates better model performance.
  + **NO2:** MSE is 346.42 for NO2, suggesting a larger spread of squared errors compared to SO2.
  + **RSPM/PM10:** The model has a higher MSE of 2972.43 for RSPM/PM10, signifying more variability in the squared errors.
* **Root Mean Squared Error (RMSE):**
  + **SO2:** The RMSE of 15.52 is the square root of the MSE for SO2, providing an interpretable scale similar to the original target variable.
  + **NO2:** RMSE is 18.60 for NO2, indicating the average magnitude of error in the model's predictions for NO2.
  + **RSPM/PM10:** The RMSE of 54.49 for RSPM/PM10 is relatively high, suggesting larger errors in predicting this variable.
* **R-squared:**
  + **SO2:** The R-squared value of 0.34 indicates that the model explains about 34% of the variability in SO2 values. A higher R-squared is desirable.
  + **NO2:** With an R-squared of 0.14, the model's explanatory power is limited for NO2.
  + **RSPM/PM10:** The negative R-squared (-1.02) suggests that the model doesn't fit well to the RSPM/PM10 data, and its predictive power is worse than a simple mean.