**AIR QUALITY ASSESSMENT IN TN**

**(DATA ANALYTICS WITH COGNOS)**

**Phase-5 Submission Document**

**Project Title: Air Q Assessment TN**

**Phase-5: Project Documentation & Submission**

**Topic:**

*In this section we will document the complete project and prepare it for the submission.*

**Steps involved in Phase-5 Documentation:**

* Describe the project’s objective,analysis approach ,visualization techniques and code implementation.
* Include example outputs of data analysis and visualizations.
* Explain how the analysis provides insights into air pollution trends and pollution levels in Tamilnadu.

**Project Definition:**

The project aims to analyse and visualize air quality data from monitoring stations in Tamil Nadu. The objective is to gain insights into air pollution trends, identify areas with high pollution levels, and develop a predictive model to estimate RSPM/PM10 levels based on SO2 and NO2 levels. This project involves defining objectives, designing the analysis approach, selecting visualization techniques, and creating a predictive model using Python and relevant libraries.

**1.Project Objective:**

* Analyse air quality data from monitoring stations in Tamil Nadu.
* Gain insights into air pollution trends.
* Identify areas with high pollution levels.
* Develop a predictive model to estimate RSPM/PM10 levels based on SO2 and NO2 levels.

**2.Analysis Approach:**

* Data Collection: Specify how you will gather air quality data, including sources, frequency, and data formats.
* Data Pre-processing: Outline steps for cleaning, handling missing data, and ensuring data quality.
* Exploratory Data Analysis (EDA): Describe techniques for exploring data to uncover trends and patterns.
* Feature Engineering: Detail how you’ll create relevant features for modelling.
* Model Selection: Decide on machine learning algorithms for building the predictive model
* Evaluation Metrics: Define how you’ll measure the model’s performance.
* Cross-validation: Plan for validating the model’s performance.
* Hyper parameter Tuning: Discuss methods for optimizing model parameters.

**3.Visualization Techniques:**

* Select appropriate visualization techniques based on the nature of the data.
* Line charts for time series analysis of air quality trends.
* Heat maps or spatial maps for pinpointing pollution hotspots.
* Scatter plots or correlation matrices to understand relationships between variables.
* Ensure that the visualizations are user-friendly and insightful for stakeholders.

**4.Python and Libraries:**

* Specify the Python libraries you plan to use for data analysis and modelling (e.g., pandas, NumPy, scikit-learn).
* Mention any specific data visualization libraries (e.g., Matplotlib, Seaborn) that will be employed.
* Include any additional libraries required for geospatial analysis if applicable.

**Steps involved in Processing data set:**

To incorporate machine learning algorithms to improve the accuracy of a predictive model, you can follow these steps:

**Data Pre-processing:**

* Collect and clean your data, handling missing values and outliers.
* Encode categorical variables and scale numerical features if necessary.

**Split the Data:**

* Divide your dataset into training, validation, and test sets.
* Select machine learning algorithms suitable for your problem (e.g., regression, classification, clustering).

**1.DESIGN THINKING AND PRESENT IN FORM OF DOCUMENT**

**1.Empathize:**

* Understand the needs and challenges of all stakeholders involved in the house price prediction process, including homebuyers, sellers,real estate professionals, appraisers, and investors.
* Conduct interviews and surveys to gather insights on what users value in property valuation and what information is most critical for their decision-making.

**2.Define:**

* Clearly articulate the problem statement, such as "How might we predict house prices more accurately and transparently using machine learning?"
* Identify the key goals and success criteria for the project, such as increasing prediction accuracy, reducing bias, or improving user trust in the valuation process.

**3.Ideate:**

* Brainstorm creative solutions and data sources that can enhance the accuracy and transparency of house price predictions.
* Encourage interdisciplinary collaboration to generate a wide range of ideas, including the use of alternative data, new algorithms, or improved visualization techniques.

**4.Prototype:**

* Create prototype machine learning models based on the ideas generated during the ideation phase.
* Test and iterate on these prototypes to determine which approaches are most promising in terms of accuracy and usability.

**5.Test:**

* Gather feedback from users and stakeholders by testing the machine learning models with real-world data and scenarios.
* Assess how well the models meet the defined goals and success criteria, and make adjustments based on user feedback.

**6.Implement:**

* Develop a production-ready machine learning solution for predicting house prices, integrating the best-performing algorithms and data sources.
* Implement transparency measures, such as model interpretability tools, to ensure users understand how predictions are generated.

**7.Evaluate:**

* Continuously monitor the performance of the machine learning model after implementation to ensure it remains accurate and relevant in a changing real estate market.
* Gather feedback and insights from users to identify areas for improvement.

**8.Iterate:**

* Apply an iterative approach to refine the machine learning model based on ongoing feedback and changing user needs.
* Continuously seek ways to enhance prediction accuracy, transparency, and user satisfaction.

**9.Scale and Deploy:**

* Once the machine learning model has been optimized and validated,deploy it at scale to serve a broader audience, such as real estate professionals, investors, and homeowners.
* Ensure the model is accessible through user-friendly interfaces and integrates seamlessly into real estate workflows.

**10.Educate and Train:**

* Provide training and educational resources to help users understand how the machine learning model works, what factors it considers,and its limitations.
* Foster a culture of data literacy among stakeholders to enhance trust in the technology.

**2.DESIGN INTO INNOVATION**

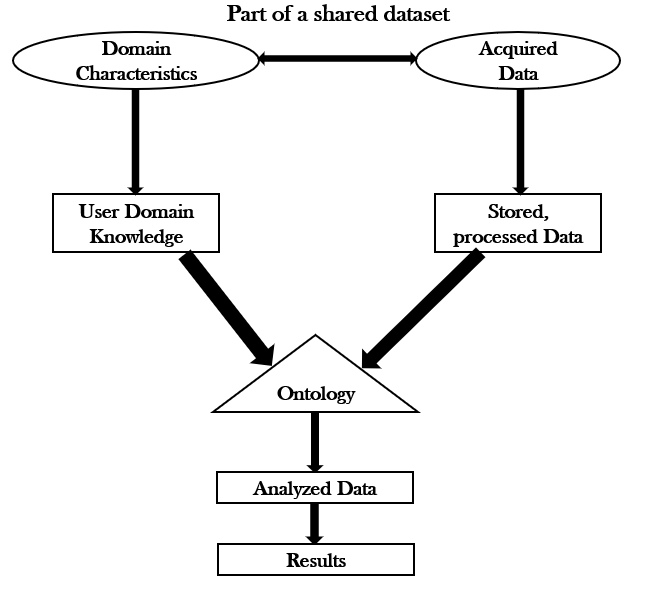
**1. Data Collection:**

Gather a comprehensive dataset that includes features such as location, size, age, amenities, nearby schools, crime rates, and other relevant variables.

**2.Data Preprocessing:**

Clean the data by handling missing values, outliers, and encoding categorical variables. Standardize or normalize numerical features as necessary.

Data preprocessing is **a step in the data mining process that cleans, transforms, and integrates data to make it ready for analysis**. The goal of data preprocessing is to improve the quality of the data and make it more suitable for the specific data mining task



**3.Feature Engineering:**

Create new features or transform existing ones to extract more valuable information. For example, you can calculate the distance to the nearest public transportation, or create a feature for the overall condition of the house.

**4.Model Selection:**

Choose the appropriate machine learning model for the task.Common models for regression problems like house price prediction include Linear Regression, Decision Trees, Random Forest, Gradient Boosting, and Neural Networks.

**5. Training:**

Split the dataset into training and testing sets to evaluate the model's performance. Consider techniques like cross-validation to prevent overfitting.

**6. Hyperparameter Tuning:**

Optimize the model's hyperparameters to improve its predictive accuracy. Techniques like grid search or random search can help with this.

**7.Evaluation Metrics:**

Select appropriate evaluation metrics for regression tasks, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE). Choose the metric that aligns with the specific objectives of your project.

**8.Regularization:**

Apply regularization techniques like L1 (Lasso) or L2 (Ridge)regularization to prevent overfitting.

**9.Feature Selection:**

Use techniques like feature importance scores or recursive feature elimination to identify the most relevant features for the prediction.

**10. Interpretability:**

Ensure that the model's predictions are interpretable and explainable. This is especially important for real estate applications where stakeholders want to understand the factors affecting predictions.

**11. Deployment:**

Develop a user-friendly interface or API for end-users to input property details and receive price predictions.

**12. Continuous Improvement:**

Implement a feedback loop for continuous model improvement based on user feedback and new data.

**13. Ethical Considerations:**

Be mindful of potential biases in the data and model. Ensure fairness and transparency in your predictions.

**14. Monitoring and Maintenance:**

Regularly monitor the model's performance in the real world and update it as needed.

**15. Innovation:**

Consider innovative approaches such as using satellite imagery or IoT data for real-time property condition monitoring, or integrating natural language processing for textual property descriptions.

**3.BUILD LOADING AND PREPROCESSING THE DATASET**

**1. Data Collection:**

Obtain a dataset that contains information about houses and their corresponding prices. This dataset can be obtained from sources like real estate websites, government records, or other reliable data providers.

**2. Load the Dataset:**

* Import relevant libraries, such as pandas for data manipulation and numpy for numerical operations.
* Load the dataset into a pandas DataFrame for easy data handling.You can use pd.read\_csv() for CSV files or other appropriate functions for different file formats.

**3. Data Exploration:**

Explore the dataset to understand its structure and contents.Check for the presence of missing values, outliers, and data types of each feature.

**4. Data Cleaning:**

Handle missing values by either removing rows with missing data or imputing values based on the nature of the data.

**5. Feature Selection:**

Identify relevant features for house price prediction. Features like the number of bedrooms, square footage, location, and amenities are often important.

We are selecting numerical features which have more than 0.50 or less than -0.50 correlation rate based on Pearson Correlation Method—which is the default value of parameter "method" in corr() function. As for selecting categorical features, select the categorical values which believe have significant effect on the target variable such as Heating and MSZoning.

**6. Feature Engineering:**

Create new features or transform existing ones to capture additional information that may impact house prices. For example, you can calculate the price per square foot.

**7. Data Encoding:**

Convert categorical variables (e.g., location) into numerical format using techniques like one-hot encoding.

**8. Train-Test Split:**

Split the dataset into training and testing sets to evaluate the machine learning model's performance.

**4.PERFORMING DIFFERENT ACTIVITIES LIKE FEATURE ENGINEERING, MODEL TRAINING,EVALUATION etc.,**

**1. Feature Engineering:**

* As mentioned earlier, feature engineering is crucial. It involves creating new features or transforming existing ones to provide meaningful information for your model.
* Extracting information from textual descriptions (e.g., presence of keywords like "pool" or "granite countertops").
* Calculating distances to key locations (e.g., schools, parks) if you have location data.

**2. Data Preprocessing & Visualisation:**

Continue data preprocessing by handling any remaining missing values or outliers based on insights from your data exploration.

**Data source:**

Dataset Link:

https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014

**Necessary steps to follow:**

* Import Libraries
* Load the Dataset
* Exploratory Data Analysis (EDA)
* Feature Engineering
* Split the Data
* Feature Scaling

**1.Import Libraries:**

Start by importing the necessary libraries for pre-processing and loading the data set of Air Q Assessment in Tamil Nadu in 2014.

**Program:**

Import pandas as pd

Import numpy as np

From sklearn.model\_selection import train\_test\_split

From sklearn.preprocessing import StandardScaler

**2.Load the Dataset:**

Load your dataset into a Pandas Dataframe. You can typically find house price datasets in CSV format, but you can adapt this code to other formats as needed.

**Program:**

df = pd.read\_csv(‘ E:\cpcb\_dly\_aq\_tamil\_nadu\_2014.csv ‘)

Pd.read()

**3.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:**

# Check for missing values

Print(df.isnull().sum())

# Explore statistics

Print(df.describe())

# Visualize the data (e.g., histograms, scatter plots, etc.)

**4.Feature Engineering:**

Depending on your dataset, you may need to create new features or transform existing ones. This can involve one-hot encoding categorical variables, handling date/time data, or scaling numerical features.

**Program:**

# Example: One-hot encoding for categorical variables

df = pd.get\_dummies(df, columns=[‘ Avg. Area Income ‘,

‘ Avg. AreaHouse Age ‘])

**5.Split the Data:**

Split your dataset into training and testing sets. This helps you evaluate your model’s performance later.

**Program:**

X = df.drop(‘price’, axis=1) # Features

Y = df[‘price’] # Target variable

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

**6.Feature Scaling:**

Apply feature scaling to normalize your data, ensuring that all features have similar scales. Standardization (scaling to mean=0 and std=1) is a common choice.

**Program:**

Scaler = StandardScaler() X\_train = scaler.fit\_transform(X\_train) X\_test = scaler.transform(X\_test)

**Importance of loading and processing dataset:**

Loading and pre-processing the dataset is an important first step in building any machine learning model. However, it is especially important for air quality assessment models.

By loading and pre-processing the dataset, we can ensure that the machine learning algorithm is able to learn from the data effectively and accurately.

To load and pre-process the air quality dataset for Tamil Nadu in 2014, we can follow these steps using Python and the Pandas library.

**Import the necessary libraries:**

import pandas as pd

**Load the dataset into a Pandas DataFrame:**

file\_path = ‘cpcb\_dly\_aq\_tamil\_nadu\_-2014.csv’

df = pd.read\_csv(file\_path)

**Explore the dataset to get an understanding of its structure and contents:**

# Display the first few rows of the dataset

Print(df.head())

# Get basic statistics of the dataset

Print(df.describe())

**Data Pre-processing:**

Depending on the quality of our dataset and our analysis goals, we may need to perform various pre-processing steps, such as handling missing values, renaming columns, and converting data types. Here are some common pre-processing tasks:

Handling missing values:

We can use methods like fillna() or dropna() to handle missing data.

Renaming columns:

we can use rename() to give more meaningful names to columns.

Converting data types:

Ensure that numerical columns are of the correct data type (e.g., float or int).

For example, to handle missing values by filling them with the mean value for each column:

# Fill missing values with the mean of each column

df = df.fillna(df.mean())

**Filter the dataset for the year 2014:**

# dataset has a ‘date’ column

df[‘date’] = pd.to\_datetime(df[‘date’])

# Convert the ‘date’ column to datetime

df\_2014 = df[df[‘date’].dt.year == 2014]

Now, we have a Pandas Data Frame df\_2014 containing the air quality data for Tamil Nadu in 2014.

Remember to adjust these steps according to the actual structure and quality of our dataset. Data pre-processing often depends on the specific characteristics of our data and the objectives of our analysis.

**How to overcome the challenges of loading and pre-processing a air quality analysis in Tamil Nadu in 2014 dataset:**

There are a number of things that can be done to overcome the challenges of loading and pre-processing a air quality analysis in Tamil Nadu in 2014 .

* **Use a data pre-processing library:**

There are a number of libraries available that can help with data pre-processing tasks, such as handling missing values, encoding categorical variables, and scaling the features.

* **Carefully consider the specific needs of your model:**

The best way to pre-process the data will depend on the specific machine learning algorithm that you are using. It is important to carefully consider the requirements of the algorithm and to pre-processes the data in a way that is compatible with the algorithm.

* **Validate the pre-processed data:**

It is important to validate the pre-processed data to ensure that it is in a format that can be used by the machine learning algorithm and that it is of high quality. This can be done by inspecting the data visually or by using statistical methods.

**1.Loading the dataset:**

* Loading the dataset using machine learning is the process of bringing the data into the machine learning environment so that it can be used to train and evaluate a model.
* The specific steps involved in loading the dataset will vary depending on the machine learning library or framework that is being used. However, there are some general steps that are common to most machine learning frameworks:

**a. Identify the dataset:**

The first step is to identify the dataset that you want to load. This dataset may be stored in a local file, in a database, or in a cloud storage service.

**b. Load the dataset:**

Once you have identified the dataset, you need to load it into the machine learning environment. This may involve using a built-in function in the machine learning library, or it may involve writing your own code.

**c. Pre-process the dataset:**

Once the dataset is loaded into the machine learning environment ,you may need to pre process it before you can start training and evaluating your model. This may involve cleaning the data, transforming the data into a suitable format, and splitting the data into training and test sets.

**Program:**

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)

import os

for dirname, \_, filenames in os.walk(‘/input/cpcb\_dly\_aq\_tamil\_nadu\_2014.csc’):

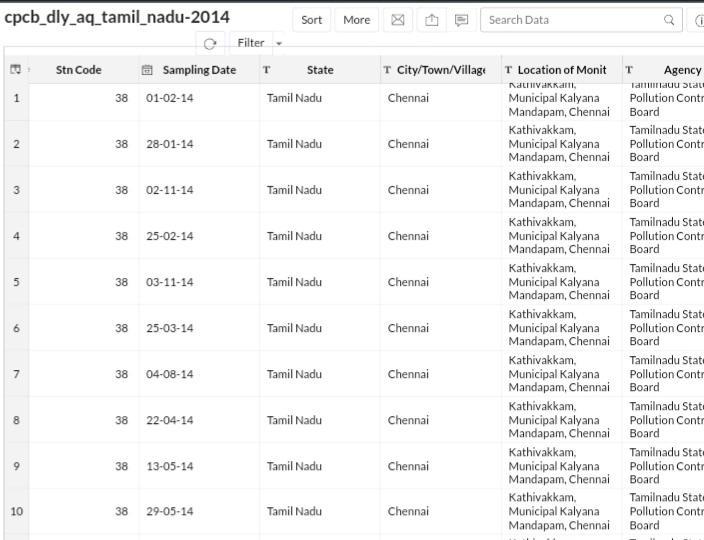
for filename in filenames:

print(os.path.join(dirname, filename))

df=pd.read\_csv(‘../input/input/cpcb\_dly\_aq\_tamil\_nadu\_2014.csc’,encoding=’cp1252’)

df.head()

**Output:**



df.shape

**Output:**

(2879, 11)

df[df[‘state’]==’Tamil Nadu’][‘location’].unique()

**Output:**

array([‘Madras’, ‘Turicorin’, ‘Tuticorin’, ‘Coimbatore’, ‘Madurai’,

‘Salem’, ‘Chennai’, ‘Thoothukudi’, ‘Trichy’, ‘Mettur’, ‘Cuddalore’],

Dtype=object)

df[‘location’]=df[‘location’].replace((‘Madras’, ‘Turicorin’, ‘Thoothukudi’, ‘Mettur’),(‘Chennai’, ‘Tuticorin’, ‘Tuticorin’, ‘Salem’))

df[df[‘state’]==’Tamil Nadu’][[‘location’, ‘rspm’]].groupby([‘location’]).agg(‘mean’).sort\_values(‘rspm’,

ascending=False).style.background\_gradient(cmap=’cool’

**Output:**

**#Let’s check which type of area causing more pollution in Tuticorin**

df[df[‘location’]==’Tuticorin’][[‘type’, ‘rspm’]].groupby([‘type’]).agg(‘mean’).sort\_values(‘rspm’,ascending=False).style.background\_gradient(cmap=’inferno ’)

**Output:**

****

**#As we know Industrial areas causing more pollution let’s check which industrial areas causing more pollution in Tamil Nadu**

df[(df[‘type’]==’industrial’)|(df[‘state’]==’Tamil Nadu’)][[‘location’,

‘rspm’]].groupby([‘location’]).agg(‘mean’).sort\_values(‘rspm’,

ascending=False).style.background\_gradient(cmap=’rainbow’)

**Output:**



**Data visualization:**

Data visualization is a crucial aspect of data analysis and communication. It involves creating graphical representations of data to help people understand patterns, trends, and insights in the data.

**Types of Visualizations:**

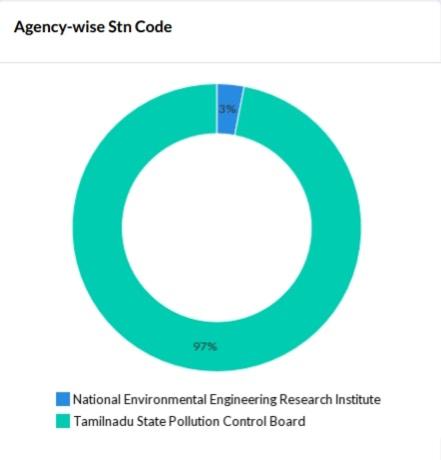
There are various types of data visualizations, including bar charts, line graphs, scatter plots, heatmaps, pie charts, histograms, and more.

The choice of visualization depends on the data and the insights you want to convey.

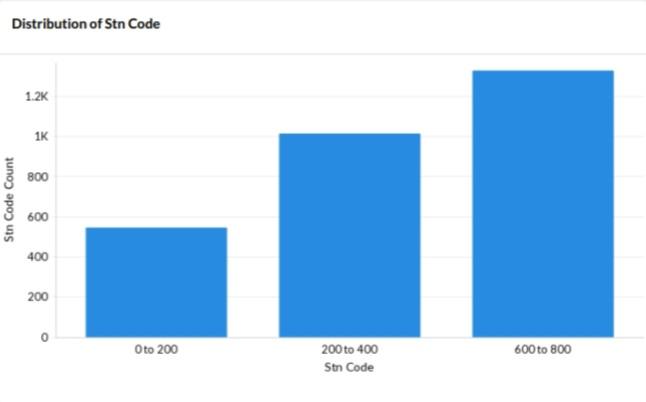
**Data Visualization using data visualization libraries (e.g., Matplotlib, Seaborn):**

* Air quality data can be visualized by combining real-time monitoring data with Python programming. Interactive graphs can be created.
* Air quality monitors are equipped with sensors that detect specific pollutants. Some monitors use lasers to scan particulate matter density in a cubic meter of air. Others use satellite imaging to measure energy reflected or emitted by the Earth.

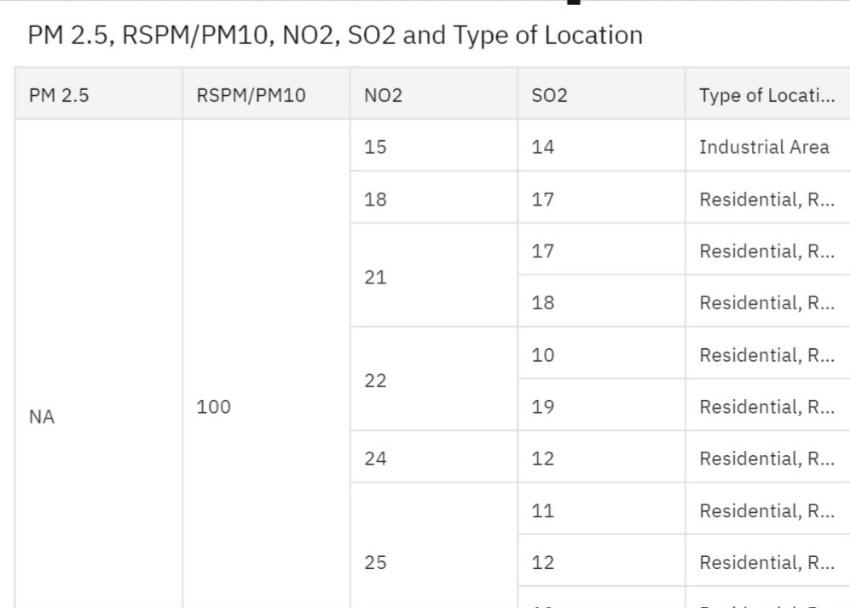
**Agency wise Stn-Code(Station-Code):**

****

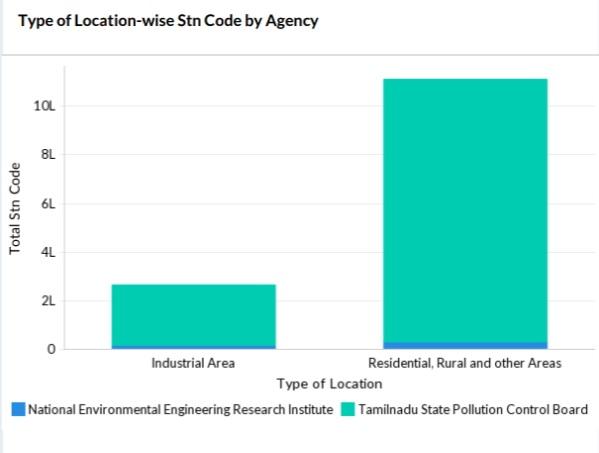
**Distribution of Stn-Code:**

****

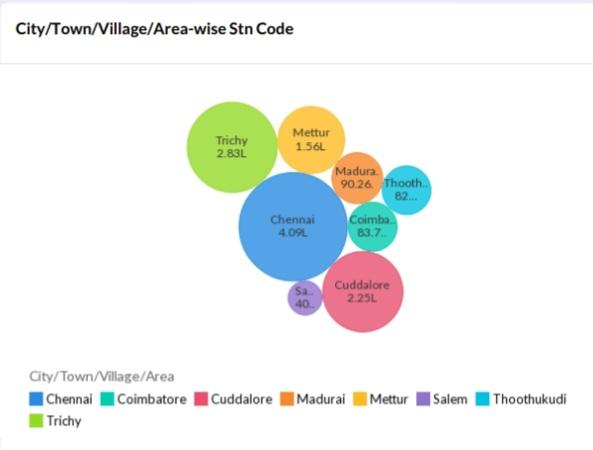
**PM 2.5,RSPM/PM10,NO2,SO2 and Type of Location:**

****

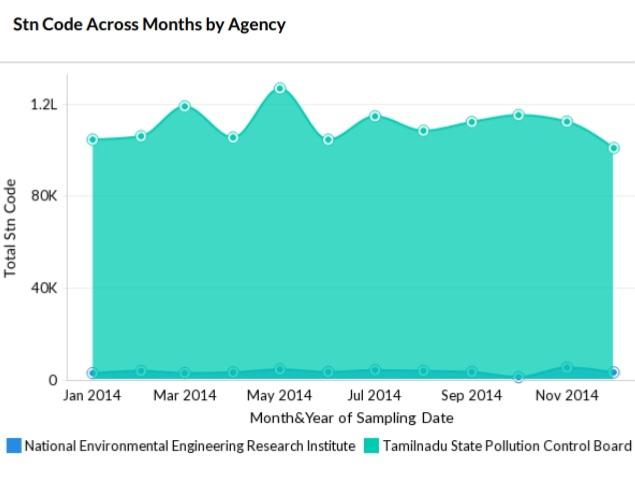
**Type of location wise Stn-Code by Agency:**

****

**City/Town/Village/Area-wise Stn-Code:**

****

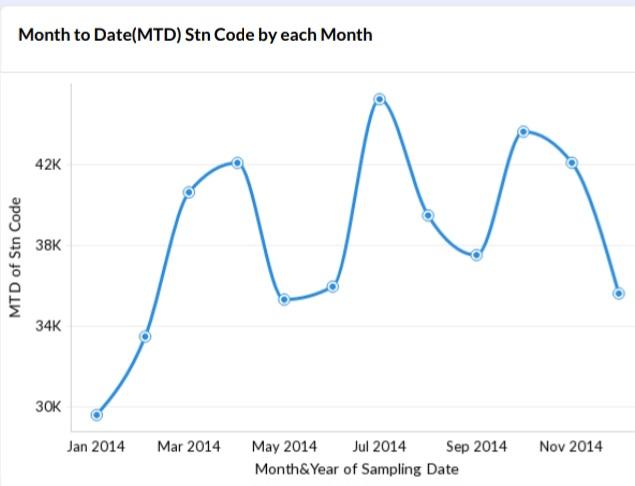
**Stn-Code Across Month by Agency:**

****

National Environmental Engineering Research Institute and Tamil Nadu state pollution control board both agencies are involved in the analysis of air quality in Tamilnadu.

****

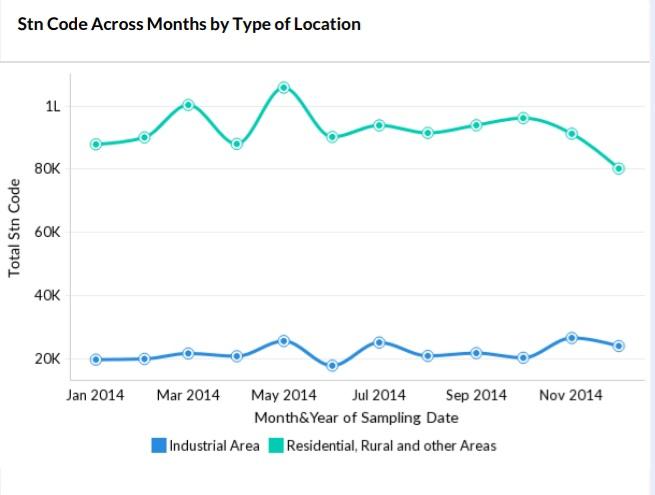
**Month to Date(MTD)Stn-Code by Each Month:**

****

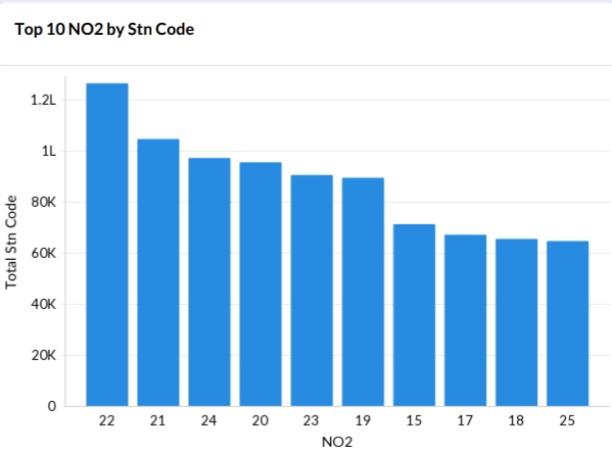
The above graph clearly explains that July month of 2014 has the maximum reach of samples in the Tamilnadu state.

**Stn-Code Across Months by Type of Location:**



Air pollution in the industrial areas ,residential and rural areas of TamilNadu in 2014

**Top 10 NO2 by Stn-Code:**



**Air Pollutant:**

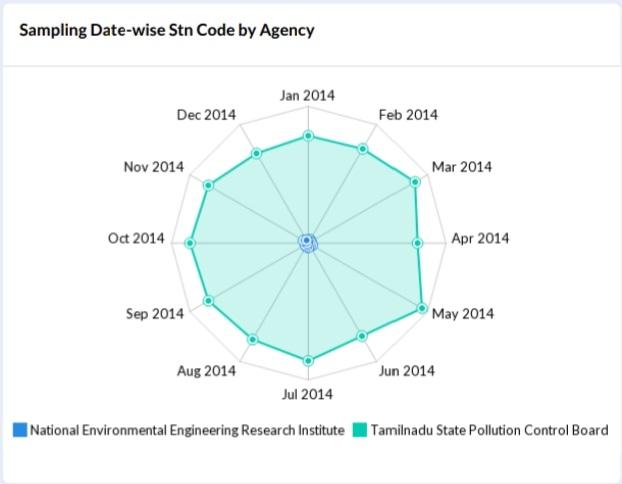
An air pollutant is **any gas or particulate that, at high enough concentration, may be harmful to life, the environment, and/or property.**

**Sources of Air Pollutants:**

* Factories
* Power plants
* Automobile exhausts
* Burning of firewood and dung cakes
* Fossil fuels such as fuel oil, gasoline, and natural gas that are burned in power plants, automobiles, and other combustion sources
* Dust generated by crushing, grinding processes
* Volcanic activities
* Industrial processes
* Production of sulphuric acid



**Sampling Data wise Stn-Code by Agency:**

****

**Stn-Code Distribution Across Months by Agency:**

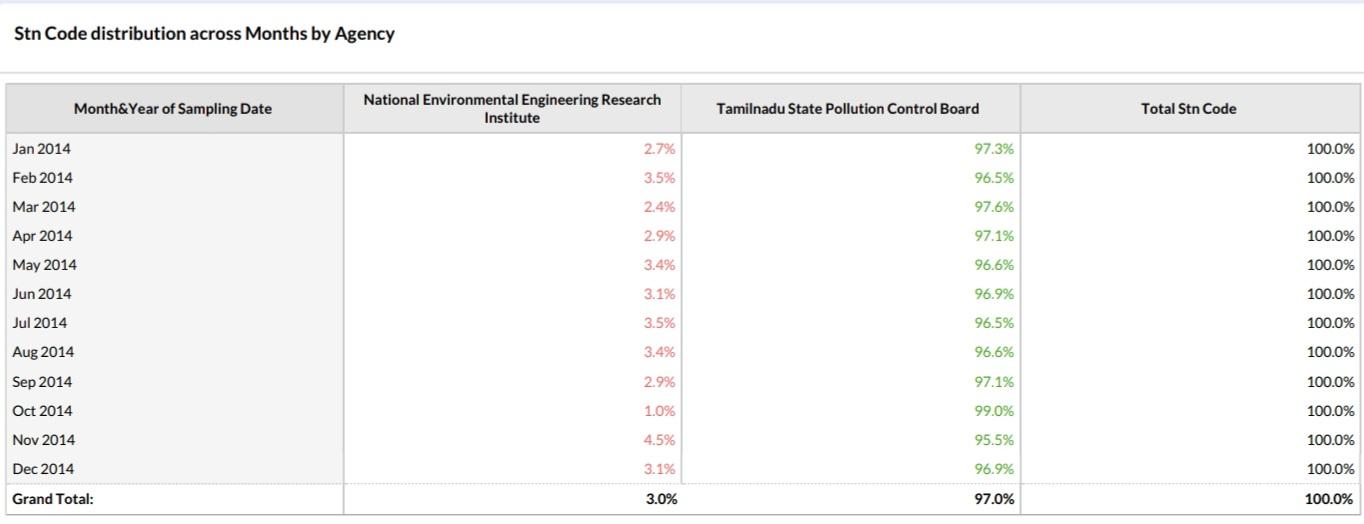
The Tamil Nadu Pollution Control Board (TNPCB) operates air quality monitoring stations in 20 districts across the state.

The TNPCB has eight stations in

* Chennai,
* Three in Thoothukudi,
* Three in Coimbatore,
* Salem.

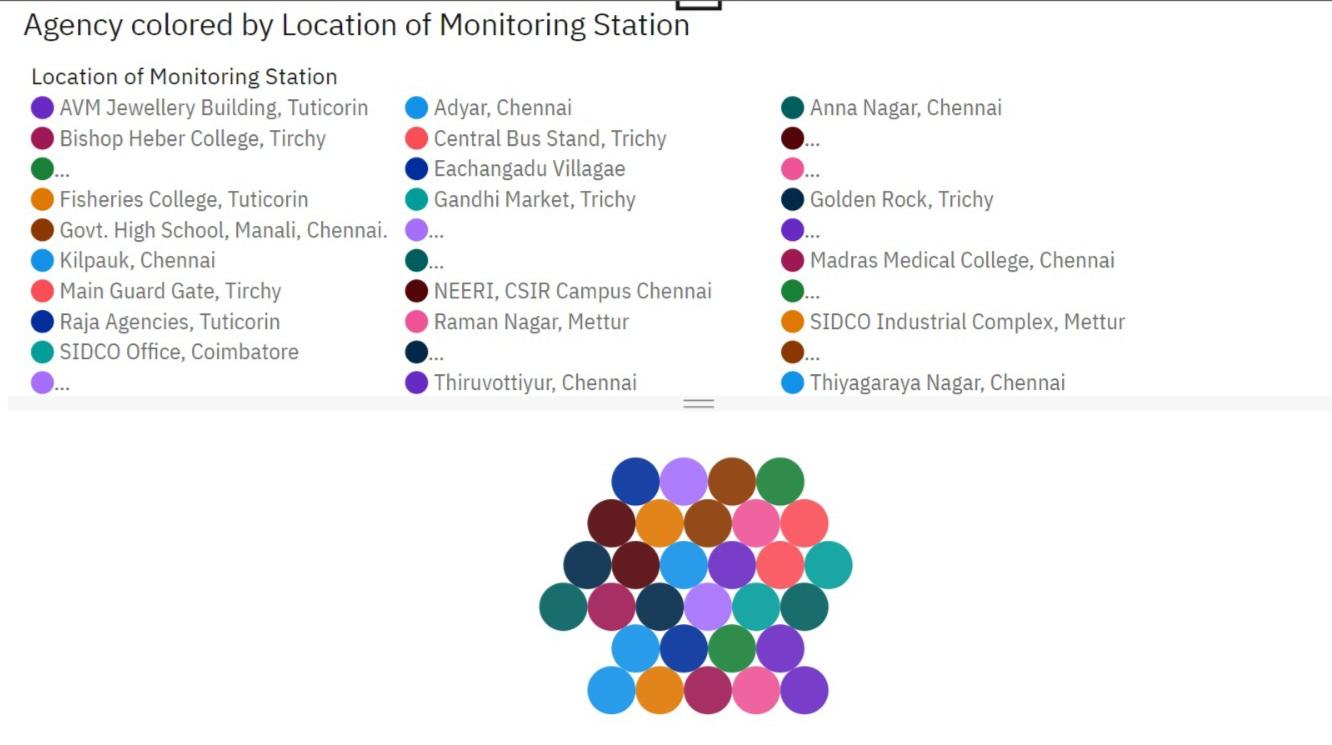
**The following agencies monitor air quality twice a week:**

* Pollution Control Committees
* Central Pollution Control Board
* National Environmental Engineering Research Institute
* State Pollution Control Boards



**Agency coloured by location of monitoring station:**

Tamil Nadu Pollution Control Board is operating eight ambient air quality monitoring stations in Chennai, Three ambient air quality monitoring stations in Thoothukudi, Three ambient air quality monitoring stations in Coimbatore ,One ambient air quality monitoring stations in Salem, Three ambient air quality monitoring stations in Madurai, Five ambient air quality monitoring stations in Trichy, Three ambient air quality monitoring stations in Cuddalore and Two ambient air quality monitoring stations in Mettur under National Air Quality Monitoring Programme (NAMP) funded by Central Pollution Control Board.



## 

**BENEFITS OF AIR QUALITY ASSESSMENT:**

**1.Improve air quality :**

Monitoring helps to identify areas with poor air quality and the pollutants responsible for it. This information can be used to implement air pollution control measures to improve air quality. Reducing the levels of pollutants in the air can lead to improved health outcomes for the population and a better quality of life.

**2.Monitor compliance with regulations:**

A[ir quality sensors](https://airly.org/en/features/air-quality-sensors/) and other devices make it possible to keep an eye on the emissions from industrial sources, such as power plants and factories, to ensure they meet the standards set by government agencies and adjust your outdoor activities accordingly. One of the main benefits of air quality monitoring is that it helps us to ensure that the air we breathe is safe.

**3.Monitor climate change:**

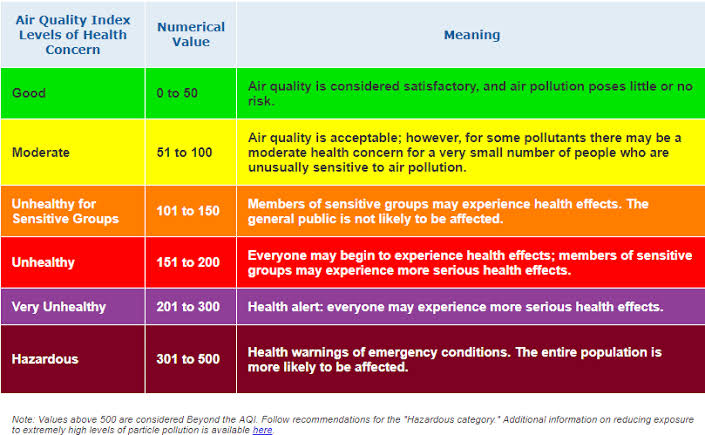
Changes in weather patterns, such as increased frequency of heat waves and wildfires, can affect the levels of pollutants in the air. By monitoring these changes, air quality monitoring can help to identify the impact of climate change on air quality and take action to mitigate it.

**4.Support research and development:**

Collected data on air quality is a unique source of inspiration for research and development of new pollution control technologies that have the potential to reduce emissions from industrial sources.

**5.Protect health:**

Why is air quality monitoring important? Pollution has been linked to a range of health problems, including respiratory and cardiovascular diseases. Air quality monitoring can help to identify areas where the air is polluted and take action to protect public health.



**AIR QUALITY MONITORING-IMPORTANCE:**

Air quality monitoring is important for many reasons, including:

* Improving air quality
* Protecting public health
* Ensuring compliance with regulations
* Identifying pollution sources
* Monitoring climate change
* Supporting research and development
* Determining if air pollution control programmes are working efficiently
* Detecting sudden increases in pollution
* Minimizing the detrimental effects of pollution
* Monitoring the presence of pollutants
* Resulting in better environmental conditions for humans to reside

Air quality monitoring is an important tool for improving air quality, protecting public health, and ensuring compliance with regulations. It can also be used to identify pollution sources, monitor climate change, or support research and development.

### Benefits of Air Quality Analysis in TN:

1.The data collected from air quality monitoring helps us assess impacts caused by poor air quality on public health.

2.Air quality data helps us determine if an area is meeting the air quality standards devised by CPCB, WHO or OSHA.

3.The data collected from air quality monitoring would primarily help us identify polluted areas, the level of pollution and air quality level.

4.Air quality monitoring would assist in determining if air pollution control programmes devised in a locality are working efficiently or not.

5.Air quality data helps us understand the mortality rate of any location due to air pollution. We can also assess and compare the short term and long term diseases/disorders which are a result of air pollution.

6.Based upon the data collected control measures can be devised for protection of the environment and health of all living organisms.

**Basically Air Quality Monitoring Benefits us by helping us protect our mother earth and our lives. Air Quality Monitoring is the first step to understand air pollution and Regular Air Quality Monitoring would give us an exact idea about the pollution level in our surroundings and help us protect ourselves from various treacherous diseases.**

**CONCLUSION OF AIR QUALITY ASSESSMENT PROJECT**

**(DATA ANALYTICS WITH COGNOS):**

From the above data analysis, we see that the majorly affected places in Tamilnadu by air pollution.The Tamil Nadu Pollution Control Board (TNPCB) operates air quality monitoring stations in 20 districts across the state.

The TNPCB has eight stations in

* Chennai
* Three in Thoothukudi
* Three in Coimbatore
* Salem

States like Chennai,Thoothukudi,Coimbatore and Salem are heavily polluted and require immediate action.

We also saw that even if a district had a high level of pollutants, there were some regions in the states that were not polluted.

In this Phase-5 documentation we described the project by data set analysis,visualization techniques and the code implementation processes.

We explained how the analyses provide insights into air pollution trends and pollution levels in Tamilnadu in 2014.

In the Phase-5 Conclusion of Air quality assessment project(Data Analytics with Cognos) we performed the air quality analysis and created visualizations and calculated average SO2, NO2, and RSPM/PM10 levels across different monitoring stations, cities, or areas.Then we identified the pollution trends and areas with high pollution levels and created visualizations using data visualization