DHAANISH AHMED COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Domain Name : Data Analytics with Cognos

Project Title : Air Quality Analysis in Tamil Nadu

phase 1: Project Definition and Design Thinking

Name of the Studen :Deva R

PROJECT DEFINITION :

The project aims to analyze 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.

Air quality analysis in Tamil Nadu is the process of monitoring and assessing the quality of the air in the state. This is done by collecting and analyzing data on air pollutants such as particulate matter (PM), sulfure dioxide (SO2), and nitrogen dioxide (NO2).

Air quality analysis in Tamil Nadu is important because air pollution can have a significant impact on human health and the environment. Air pollution can cause respiratory problems, heart disease, and cancer. It can also damage plants and animals , and contribute to climate change.

The Tamil Nadu Pollution Control Board (TNPCB) is responsible for monitoring and assessing air quality in the state. The TNPCB operates a network of air quality monitoring stations across Tamil Nadu. These stations collect data on air pollutants on a continuous basis.

The TNPCB uses the data collected from its air quality monitoring stations to calculate the Air Quality Index (AQI). The AQI is a measure of the overall air quality in a particular area. It is calculated using a formula that takes into account the concentrations of different air pollutants.

The AQI is divided into six categories: good, satisfactory, moderate, poor, very poor, and severe. The higher the AQI, the worse the air quality.

- Clearly state your project's objectives. For example:

- Analyze historical air quality data to identify long-term trends and seasonal variations in air pollution levels.

- Identify regions in Tamil Nadu with consistently high levels of air pollution.

- Develop a predictive model to estimate RSPM/PM10 levels based on SO2 and NO2 levels to help in early warning and intervention

DESIGN THINKING :

Design thinking is a non-linear, iterative process that teams use to understand users, challenge assumptions, redefine problems, and create innovative solutions to prototype and test. Design thinking is a human-centered approach to design that focuses on the needs and wants of the users.

Design thinking can be applied to air quality analysis in data analysis in the following ways:

1. **Empathize with the users.** Who are the users of the air quality analysis? What are their needs and concerns? What do they hope to achieve by using the data?

2. **Define the problem.** What is the specific problem that the air quality analysis is trying to solve? What are the key factors that contribute to the problem?
3. **Ideate.** Brainstorm a variety of possible solutions to the problem. Be creative and don't be afraid to think outside the box.

4. **Prototype**. Build prototypes of the different solutions and test them with users. Get feedback from users and refine the prototypes as needed.

5. **Test and implement.** Once the prototype has been refined, test it in a real-world setting. Collect data and feedback from users, and make further adjustments as needed.

Here are some specific examples of how design thinking can be applied to air quality analysis in data analysis:

- **Developing a new air quality monitoring system.** A team could use design thinking to develop a new air quality monitoring system that is more accurate, affordable, and easier to use than existing systems.
- **Creating a data visualization tool for air quality data**. A team could use design thinking to develop a data visualization tool that makes it easy for users to understand and interpret air quality data.
- **Designing a public awareness campaign about air quality.** A team could use design thinking to develop a public awareness campaign that educates people about the importance of air quality and the steps they can take to improve it.

Design thinking is a powerful tool that can be used to develop innovative and effective solutions to air quality problems. By focusing on the needs of the users and iteratively developing and testing solutions, design teams can create solutions that have a real impact on people's lives.

1,Data Collection and Preparation:

- Gather air quality data from monitoring stations in Tamil Nadu. This data can typically be obtained from government agencies or environmental organizations.
- Ensure that your data is clean, consistent, and properly formatted.
- Merge data from different monitoring stations if necessary.

2. Exploratory Data Analysis (EDA):

- Use Python libraries like Pandas, NumPy, and Matplotlib/Seaborn to perform EDA.
- Calculate statistics, histograms, and scatter plots to understand the distribution and relationships between different air pollutants (SO2, NO2, RSPM/PM10).
- Identify outliers and missing values and decide how to handle them.

3.Data Visualization:

- Choose appropriate visualization techniques to represent your findings effectively.

- Consider using libraries like Matplotlib, Seaborn, Plotly, or Bokeh for creating plots and dashboards.

- Visualize trends in air pollution over time, geographical distribution, and correlations between pollutants.

4.Spatial Analysis:

- Use geospatial libraries like GeoPandas or Folium to create maps showing pollution levels across Tamil Nadu.

- Identify areas with consistently high pollution levels.

5. Predictive Modeling:

- Split your dataset into training and testing sets.

- Choose a machine learning algorithm suitable for regression tasks (e.g., Linear Regression, Random Forest, XGBoost).

- Train your model on historical data with SO2 and NO2 levels as features and RSPM/PM10 levels as the target variable.

- Evaluate the model's performance using metrics like Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE).

- Fine-tune your model, if necessary, to improve its accuracy.

7. Model Interpretation:

- Interpret the model to understand the importance of SO2 and NO2 levels in predicting RSPM/PM10 levels.
- Use feature importance plots to visualize this.

8. Communication of Results:

- Create a report or presentation summarizing your findings, insights, and the predictive model's performance.
- Share visualizations and key takeaways to make your findings accessible to a wider audience, including policymakers and the public.

9. Future Work and recommendations:

- Suggest potential actions or policies based on your findings to mitigate air pollution in high-risk areas.
- Identify areas for further research or data collection to enhance the accuracy of your predictive model.

10. Documentation and Code Sharing:

Maintain clear and well-documented code, making it easy for others to understand and reproduce your analysis.

- Remember to stay up-to-date with the latest research and methodologies in air quality analysis and predictive modelling as you work on your project. Additionally, consider ethical considerations related to data privacy and the potential impacts of your findings on the community and the environment.

DATA SOURCE :

Stn Code	Sampling D	State	City/Town,	Location o	Agency	Type of Lo	SO2	NO2	RSPM/PM1	PM 2.5
38	01-02-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	11	17	55	NA
38	01-07-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	13	17	45	NA
38	21-01-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	12	18	50	NA
38	23-01-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	15	16	46	NA
38	28-01-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	13	14	42	NA
38	30-01-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	18	43	NA
38	02-04-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	12	17	51	NA
38	02-06-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	13	16	46	NA
38	02-11-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	10	19	50	NA
38	13-02-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	15	14	48	NA
38	18-02-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	16	32	NA
38	20-02-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	14	29	NA
38	25-02-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	13	17	17	NA
38	27-02-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	15	16	44	NA
38	03-04-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	12	17	25	NA
38	03-06-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	13	16	29	NA
38	03-11-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	11	18	29	NA
38	13-03-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	15	16	41	NA
38	18-03-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	17	43	NA
38	20-03-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	14	42	NA
38	25-03-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	17	54	NA
38	27-03-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	15	19	62	NA
38	04-01-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	15	66	NA
38	04-03-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	11	16	40	NA
38	04-08-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	17	56	NA
38	04-10-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	15	17	50	NA
38	15-04-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	12	14	49	NA
38	17-04-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	15	16	63	NA
38	22-04-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	15	18	42	NA
38	29-04-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	12	18	44	NA
38	05-06-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	13	13	43	NA
38	05-08-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	14	48	NA
38	13-05-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	14	13	63	NA
38	15-05-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	12	15	119	NA
38	20-05-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	13	18	61	NA
38	27-05-14	Tamil Nadu	Chennai	Kathivakka	Tamilnadu	Industrial A	12	16	48	NA

BENEFITS OF DATA ANALYSTICS ENABLED AIR QUALITY ANALYSIS IN TAMIL NADU:

Data analytics enabled air quality analysis can provide a number of benefits to the state of Tamil Nadu, including:

- **Improved understanding of air quality:** Data analytics can be used to identify trends and patterns in air quality data, which can help to improve our understanding of the factors that contribute to air pollution and the impact that it is having on human health and the environment.
- **More effective air quality management:** Data analytics can be used to develop and implement more effective air quality management strategies. For example, data can be used to identify areas where air pollution is highest and to target interventions to those areas.
- **Better informed decision-making:** Data analytics can help policymakers to make better informed decisions about air quality management. For example, data can be used to assess the impact of different air quality policies and to identify the most effective ways to reduce air pollution.

Here are some specific examples of how data analytics enabled air quality analysis has been used in Tamil Nadu:

- The Tamil Nadu Pollution Control Board (TNPCB) is using data analytics to track air quality data from across the state. This data is being used to identify areas where air pollution is highest and to develop targeted interventions.
- The Indian Institute of Technology Madras (IIT Madras) is developing a data analytics platform for air quality analysis. This platform will be used to collect and analyze air quality data from a variety of sources, including ground-based monitoring stations, satellite data, and social media. The platform will be used to generate real-time air quality maps and forecasts, and to identify air pollution hotspots.
- The Chennai Metropolitan Development Authority (CMDA) is using data analytics to develop a traffic management plan that aims to reduce air pollution in the city. The plan uses data from traffic sensors and air quality monitoring stations to identify congested areas and to develop strategies to improve traffic flow.

These are just a few examples of how data analytics enabled air quality analysis is being used in Tamil Nadu to improve air quality and protect public health. As data analytics technologies continue to develop, we can expect to see even more innovative and effective ways to use data to improve air quality in the state.

In addition to the benefits listed above, data analytics enabled air quality analysis can also help to:

- **Reduce the cost of air quality management:** By using data to identify and target interventions, air quality managers can reduce the cost of managing air pollution.
- **Improve public awareness of air quality:** Data analytics can be used to generate real-time air quality maps and forecasts, which can help to raise public awareness of air quality and the steps that people can take to protect themselves from air pollution.
- **Promote sustainable development:** By improving air quality, data analytics can help to promote sustainable development in Tamil Nadu.

CONCLUSION:

Here by I am concluding my knowledge at
phase 1: Project Definition and Design Thinking.