DATA ANALYTICS WITH COGNOS – DAC101

PHASE 1: PROBLEM DEFINITION AND

DESIGN THINKING

PROBLEM 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.

DESIGN THINKING:

1.PROJECT OBJECTIVES:

1.**Analyzing Air Quality Trends:** This objective involves examining historical data on air quality parameters such as particulate matter (e.g., PM2.5 and PM10), gases (e.g., nitrogen dioxide, sulfur dioxide), and meteorological factors (e.g., temperature, wind speed) to identify patterns, changes, or trends over time. This analysis helps in understanding how air quality has evolved and may provide insights into the causes and effects of air pollution.

2. **Identifying Pollution Hotspots:** Identifying pollution hotspots refers to the process of pinpointing specific geographic areas where air pollution levels consistently exceed acceptable standards or where pollution events are more frequent and severe. This objective aims to spatially locate areas with higher pollution concentrations, which can be valuable for targeted interventions and resource allocation to mitigate air pollution.

3.**Building a Predictive Model for RSPM/PM10 Levels:** This objective involves developing a mathematical or statistical model that can forecast or predict the levels of Respirable Suspended Particulate Matter (RSPM) or Particulate Matter with a diameter of 10 micrometers or less (PM10) in the air. Such a predictive model typically relies on historical air quality data, meteorological variables, and potentially other relevant factors (e.g., traffic volume, industrial activity) to estimate future pollution levels. This model can assist in early warning systems and policy planning to mitigate air pollution.

2.ANALYSIS APPROACH:

**1. Data Collection:**

* The data needed for the air quality analysis in Tamilnadu should be collected. The data is collected from the dataset link given:

DATASET LINK:[**https://tn.data.gov.in/resource/location-wise-daily-ambient-air- quality-tamil-nadu-year-2014**](https://tn.data.gov.in/resource/location-wise-daily-ambient-air-%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20%20quality-tamil-nadu-year-2014)

**2. Data Preprocessing:**

* Handle missing data: Identify and fill in missing values using interpolation or other appropriate methods.
* Data cleaning: Remove outliers and erroneous data points that may skew analysis results.
* Standardize units: Ensure that all data are in consistent units (e.g., convert different units of concentration to a common unit).
* Data normalization: Normalize the data if necessary to scale all variables to a common range.

**3. Exploratory Data Analysis (EDA):**

* Generate summary statistics and descriptive visualizations to gain initial insights into the data's distribution and characteristics.
* Plot time series graphs to visualize trends and patterns in air quality parameters over time.
* Calculate correlation matrices to identify relationships between different variables (e.g., how weather factors correlate with pollutant levels).

**4. Trend Analysis:**

* Use time series analysis techniques (e.g., moving averages, exponential smoothing) to identify long-term trends in air quality parameters.
* Explore seasonality and periodic patterns, if any, in the data.

**5. Hotspot Identification:**

* Create spatial visualizations (e.g., heatmaps, choropleth maps) to pinpoint pollution hotspots based on geographical data.
* Analyze temporal patterns in hotspots to identify recurring pollution events.

**6. Predictive Modeling:**

* Split the data into training and testing sets for model development and evaluation.
* Build predictive models (e.g., regression, time series forecasting, machine learning models) to predict future air quality levels.
* Evaluate model performance using appropriate metrics (e.g., Mean Absolute Error, R-squared) and cross-validation techniques.

**7. Visualization of Results:**

* Create informative and visually appealing charts and graphs to present analysis results, trends, and model predictions.
* Use geographic information systems (GIS) tools to visualize pollution hotspots on maps.
* Provide interactive visualizations if possible, enabling users to explore the data and findings.

**8. Interpretation and Reporting:**

* Interpret the analysis results and draw actionable insights.
* Prepare a comprehensive report summarizing the findings, including trends, hotspot locations, and predictive model outcomes.
* Communicate the results to relevant stakeholders, policymakers, and the public.

**9. Continuous Monitoring and Updates:**

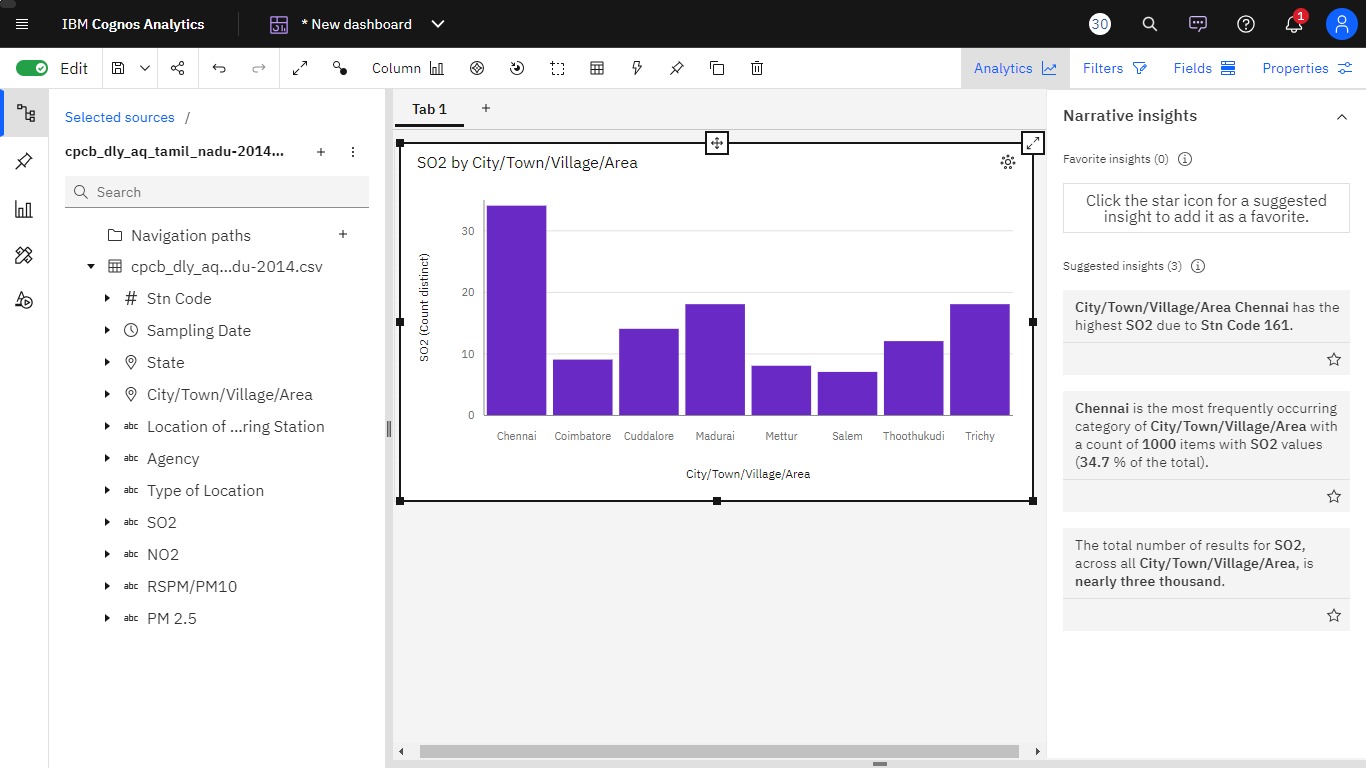
* Implement a system for ongoing data collection and analysis to monitor air quality changes over time.
* Periodically update the analysis to incorporate new data and refine predictive models.

VISUALIZATION SELECTION:

1. **Time Series Line Charts:**
   * Use line charts to display trends in air quality parameters (e.g., PM2.5, PM10) over time.
   * Line charts are effective for showing how pollution levels change throughout the day, week, or year.
2. **Heatmaps:**
   * Create heatmaps to visualize the spatial distribution of pollution levels across a geographic area.
   * Color intensity can represent pollutant concentrations, with darker colors indicating higher levels.
   * Heatmaps are excellent for identifying pollution hotspots and patterns.
3. **Choropleth Maps:**
   * Represent air quality data on a map using choropleth maps, where regions or areas are shaded or colored based on pollutant levels.
   * Choropleth maps help visualize spatial variations in pollution levels.
4. **Histograms and Bar Charts:**
   * Use histograms or bar charts to show the distribution of pollutant concentrations.
   * These charts provide insights into the frequency and range of pollution levels.
5. **Box Plots:**
   * Box plots are useful for displaying the distribution of pollutant data, including median values and outliers.
   * They can help identify the spread and central tendencies of the data.
6. **Scatter Plots:**
   * Create scatter plots to explore relationships between air quality parameters (e.g., PM2.5) and other variables like temperature or wind speed.
   * Scatter plots can reveal correlations or trends between variables.
7. **Stacked Area Charts:**
   * Use stacked area charts to visualize the composition of pollutant concentrations over time.
   * This technique shows how different pollutants contribute to overall air quality.
8. **Wind Rose Charts:**
   * Wind rose charts can help illustrate the direction and frequency of wind patterns in relation to pollution levels.
   * They are valuable for understanding how wind affects air quality.
9. **Animated Time Series Plots:**
   * Create animated plots to show changes in air quality parameters over time, allowing viewers to observe trends dynamically.
   * Animation can highlight temporal patterns and variations.
10. **Interactive Dashboards:**
    * Develop interactive dashboards that enable users to explore air quality data interactively.
    * Include filters, sliders, and dropdown menus to customize views by date, location, or pollutant.
11. **3D Visualizations (if applicable):**
    * For three-dimensional data (e.g., pollution levels in three-dimensional space), consider 3D visualizations or contour plots.
12. **Radar Charts or Spider Charts:**
    * Use radar charts to compare air quality across multiple dimensions (e.g., different pollutants) on a single chart.
13. **Word Clouds (for textual data):**
    * If you have qualitative data related to air quality (e.g., public comments), use word clouds to highlight common themes or keywords.

When choosing visualization techniques, consider the nature of your data, your audience, and the specific insights you want to convey. It's often beneficial to use a combination of visualizations to provide a comprehensive view of air quality trends and pollution levels. Additionally, ensure that your visualizations are clear, well-labeled, and easy to interpret to maximize their impact.

Top of Form



The above visualization shows the SO2 pollution in Tamilnadu. Here , the visualization technique barchart is used based on the given data.