AIR QUALITY ANALYSIS IN TAMILNADU

Phase 4: Development Part 2

TOPIC: Continue building your project by performing different analysis model building and evaluation. Perform different analysis and visualization using IBM Cognos.

INTRODUCTION

- Air quality analysis is a critical endeavor that plays a vital role in safeguarding public health, protecting the environment, and promoting sustainable urban development. In recent years, the issue of air quality has gained increasing attention due to its significant impact on the well-being of individuals and communities. Pollutants in the air, such as particulate matter (PM), volatile organic compounds (VOCs), nitrogen dioxide (NO2), sulfur dioxide (SO2), and ozone (O3), can have adverse effects on respiratory health, cardiovascular health, and overall quality of life.

- This technical project aims to address the pressing need for comprehensive air quality analysis in our region. It seeks to utilize advanced monitoring, data collection, and analytical techniques to assess air quality, identify pollution sources, and recommend mitigation strategies.

- The project will create visually appealing and informative data visualizations to make complex air quality information accessible to a wide audience. These visualizations will be used to generate reports that can be easily understood by policymakers, environmental agencies, and the general public.

- By embarking on this technical project, we aim to contribute to a healthier, more sustainable future for our community. Through the utilization of advanced technology, scientific rigor, and community involvement, we will work towards a cleaner environment and improved public health by mitigating the adverse effects of air pollution. This project aligns with the global movement towards cleaner and more sustainable cities, where the well-being of residents and the environment take center stage.

Given Dataset:

STN CODE	SAMPLING DATE	STATE	CITY	MONITORING STATION	AGENCY	TYPE OF LOCATION	SO2	NO2	RSPM/PM10	PM2.5
38	02-01-2014	TAMILNADU	CHENNAI	KATHIVAKKAM	TAMILNADU STATE POLLUTION CONTROL BOARD	INDUSTRIAL AREA	11	17	55	NA
38	07-01-2014	TAMILNADU	CHENNAI	KATHIVAKKAM	TAMILNADU STATE POLLUTION CONTROL BOARD	INDUSTRIAL AREA	13	17	45	NA
38	21-01-2014	TAMILNADU	CHENNAI	KATHIVAKKAM	TAMILNADU STATE POLLUTION CONTROL BOARD	INDUSTRIAL AREA	12	18	50	NA
38	23-01-2014	TAMILNADU	CHENNAI	KATHIVAKKAM	TAMILNADU STATE POLLUTION CONTROL BOARD	INDUSTRIAL AREA	15	16	46	NA
38	28-01-2014	TAMILNADU	CHENNAI	KATHIVAKKAM	TAMILNADU STATE POLLUTION CONTROL BOARD	INDUSTRIAL AREA	13	14	42	NA
....	....	....	....	....	....	....	....	....	....	....
773	03-12-2014	TAMILNADU	TRICHY	CENTRAL BUS STAND	TAMILNADU STATE POLLUTION CONTROL BOARD	RESIDENTIAL ,RURAL AND OTHER AREAS	15	18	102	NA
773	10-12-2014	TAMILNADU	TRICHY	CENTRAL BUS STAND	TAMILNADU STATE POLLUTION CONTROL BOARD	RESIDENTIAL ,RURAL AND OTHER AREAS	12	14	91	NA
773	17-12-2014	TAMILNADU	TRICHY	CENTRAL BUS STAND	TAMILNADU STATE POLLUTION CONTROL BOARD	RESIDENTIAL ,RURAL AND OTHER AREAS	19	22	100	NA
773	24-12-2014	TAMILNADU	TRICHY	CENTRAL BUS STAND	TAMILNADU STATE POLLUTION CONTROL BOARD	RESIDENTIAL ,RURAL AND OTHER AREAS	15	17	95	NA
773	31-12-2014	TAMILNADU	TRICHY	CENTRAL BUS STAND	TAMILNADU STATE POLLUTION CONTROL BOARD	RESIDENTIAL ,RURAL AND OTHER AREAS	14	16	94	NA

The above is the given dataset of Air Quality Analysis in Tamil Nadu in 2014. This dataset consists of 11 columns and 2880 rows.

DIFFERENT ANALYSIS

Air quality analysis in Tamil Nadu, like in any region, involves a range of analytical approaches to assess and understand air pollution and its impacts. Here are several different types of analyses commonly conducted for air quality analysis in Tamil Nadu:

1.Descriptive Analysis:

- Overview of historical air quality data to identify trends and patterns in pollutant concentrations.
- Seasonal variations in air quality, such as higher pollution levels during specific months.
- Spatial distribution of pollutants to identify pollution hotspots.

2.Time Series Analysis:

- Examination of hourly, daily, or monthly time series data to identify daily and seasonal patterns.
- Detection of long-term trends in pollutant concentrations, which may reveal improvements or deteriorations in air quality over the years.

3.Source Apportionment Analysis:

- Identifying and quantifying the contributions of various pollution sources, such as industrial emissions, vehicular traffic, and natural sources (e.g., dust storms or wildfires).
- Using techniques like chemical mass balance, receptor modeling, and dispersion modeling.

4.Air Quality Index (AQI) Calculation:

- Calculating the AQI based on measured pollutant concentrations to communicate air quality levels to the public.
- Assigning health-based categories (e.g., Good, Moderate, Unhealthy) to different AQI ranges.

5.Health Impact Assessment:

- Estimating the health impacts of poor air quality on the population, such as the number of premature deaths, hospital admissions, or exacerbation of respiratory diseases.
- Assessing the impact on vulnerable groups, including children and the elderly.

6.Regulatory Compliance Analysis:

- Evaluating whether air quality measurements comply with local and national environmental regulations and standards.
- Identifying areas that are not in compliance and potential consequences.

7.Pollutant Correlation Analysis:

- Exploring correlations between different pollutants to understand how they interact and influence each other.
- Investigating the role of secondary pollutants, which are formed through chemical reactions in the atmosphere.

8.Meteorological Influences:

- Analyzing the impact of meteorological conditions on air quality, such as the role of wind speed and direction in pollutant dispersion.
- Assessing temperature inversions and their impact on pollutant trapping.

9.Geospatial Mapping and GIS Analysis:

- Creating maps to visualize air quality data, pollutant concentrations, and pollution sources.
- Overlaying air quality data with demographic, land use, and health data to identify vulnerable communities.

10.Forecasting and Early Warning Systems:

- Developing models to predict future air quality conditions.
- Providing early warning systems to alert the public and authorities when pollution levels are expected to exceed safe limits.

11.Emission Inventory Analysis:

- Developing comprehensive emission inventories for the region to understand the sources and amounts of pollutants released into the atmosphere.

12.Policy and Mitigation Strategies:

- Evaluating the effectiveness of existing policies and regulations in reducing air pollution.
- Recommending mitigation strategies and policy changes to improve air quality.

Each of these analyses contributes to a comprehensive understanding of air quality in Tamil Nadu and supports decision-making for better air quality management and public health protection. The choice of analysis depends on the specific goals and objectives of air quality monitoring and research in the region.

VISUALIZATION

Creating visualizations for air quality analysis in Tamil Nadu can be a powerful way to communicate complex data and insights to a broader audience. Here are some common types of visualizations that can be used for air quality analysis in the region:

1.Time Series Plots:

- Line charts showing pollutant concentrations over time (e.g., hourly, daily, monthly).
- Multiple pollutants can be displayed on the same graph for comparison.
- Highlighting seasonal trends and variations.

2.Air Quality Index (AQI) Heatmap:

- A heatmap representing AQI values across different monitoring locations in Tamil Nadu.
- Color-coding for different AQI categories, making it easy to identify areas with varying air quality.

3.Spatial Distribution Maps:

- GIS maps showing the geographic distribution of air quality monitoring stations in the region.
- Using color-coded markers to indicate air quality levels at each station.

4.Pollutant Concentration Maps:

- Maps displaying spatial variations in specific pollutant concentrations across Tamil Nadu.
- Contour or gradient shading can be used to represent pollutant levels.

5.Source Apportionment Pie Charts:

- Pie charts illustrating the contributions of different pollution sources (e.g., industrial, vehicular, natural) to overall pollution levels.
- Providing insights into the sources driving poor air quality.

6.Box Plots:

- Box-and-whisker plots to visualize the distribution of pollutant concentrations, including median, quartiles, and outliers.
- Can be used to compare different pollutants or monitoring locations.

7.Correlation Matrix:

- A matrix of scatterplots showing correlations between different pollutants.
- Identifying relationships between pollutants and potential sources of secondary pollution.

8.Meteorological Influences:

- Wind rose diagrams to illustrate prevailing wind directions.
- Scatterplots or line charts showing how meteorological factors like wind speed and temperature relate to pollutant levels.

9.Emission Inventory Bar Charts:

- Bar charts or stacked bar charts representing the emissions of different pollutants by source category.
- Comparing emissions from industries, transportation, and other sectors.

10.Population Heatmaps:

- Overlaying air quality data with population density maps to identify areas with high pollution levels and high population exposure.

11.Health Impact Assessment Graphs:

- Bar charts or pie charts showing the estimated health impacts of poor air quality, such as premature deaths, hospital admissions, and respiratory problems.

12.Forecasting Graphs:

- Line charts displaying predicted air quality levels based on forecasting models.
- Issuing warnings when pollution levels are expected to exceed safety thresholds.

13.Policy Effectiveness Charts:

- Time series plots comparing air quality data before and after the implementation of specific policies or regulations.

14.Dashboard Visualizations:

- Combining multiple visual elements, such as maps, charts, and tables, into a comprehensive dashboard for a holistic view of air quality in Tamil Nadu.

These visualizations can be created using data visualization tools like Tableau, Power BI, or custom programming in languages like Python (using libraries like Matplotlib, Seaborn, and Folium) and R. The choice of visualization depends on the specific objectives of the air quality analysis and the audience you want to communicate with. Effective visualizations can help stakeholders, policymakers, and the public better understand the state of air quality in Tamil Nadu and support decision-making for cleaner and healthier air.

VISUALIZATION TYPES VERSION 11.0

IBM® Cognos® Analytics - Reporting provids many types of visualizatons for presenting your data in a way that is meaningful for your users.

You can select from a variety of chart types (such as pie, bar, line, gauge, scatter, and so on) and you can select from a variety of chart configurations (such as stacked columns, 3-D pies). Combination charts allow you to use more than one chart type within your chart.

AREA CHART:

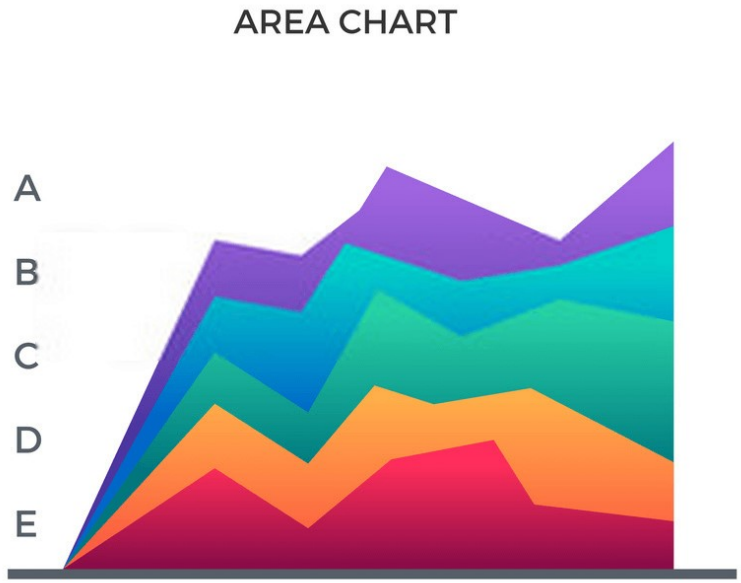
An area chart, also known as an area graph or area plot, is a type of data visualization used to represent quantitative data over a continuous interval or time period. It is a variation of a line chart where the area below the lines is filled with colors, creating a visual representation of cumulative data or the composition of multiple variables.

PROGRAM:

```
import matplotlib.pyplot as plt
import pandas as pd

# Air quality dataset
data = {
   'Date': ['2023-01-01', '2023-01-02', '2023-01-03', '2023-01-04', '2023-
01-05'],
   'PM2.5': [25, 30, 40, 20, 35],
   'PM10': [45, 50, 60, 30, 55],
}

# Create a DataFrame from the data
df = pd.DataFrame(data)
df['Date'] = pd.to_datetime(df['Date'])

# Set the Date column as the index
df.set_index('Date', inplace=True)

# Create an area chart
plt.figure(figsize=(10, 6))

plt.fill_between(df.index, df['PM2.5'], label='PM2.5', alpha=0.5)
plt.fill_between(df.index, df['PM10'], label='PM10', alpha=0.5)

# Customize the chart
plt.title('Air Quality Over Time')
plt.xlabel('Date')
plt.ylabel('Pollutant Concentration (µg/m³)')
plt.legend(loc='best')
```

```
# Rotate x-axis labels for better readability
plt.xticks(rotation=45)

# Show the chart
plt.tight_layout()
plt.show()
```

OUTPUT:



BOX LOT:

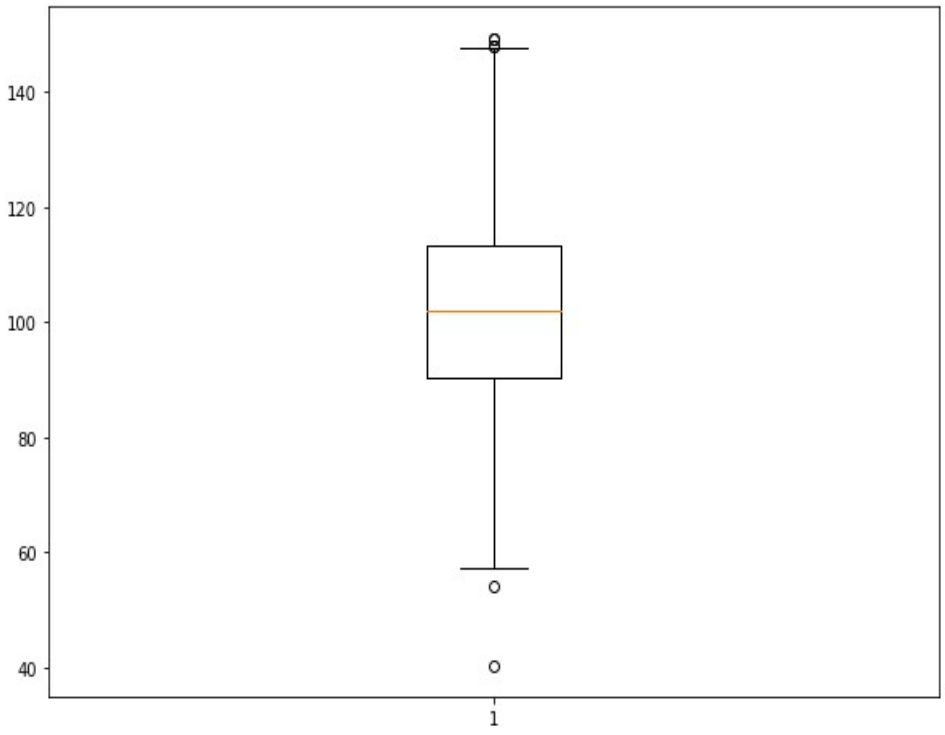
A box plot, also known as a box-and-whisker plot, is a graphical representation of the distribution of a dataset. It provides a summary of key statistical measures and visualizes the spread and central tendency of the data. You can use box plots for identifying outliers and for comparing distributions.

PROGRAM:

```
import matplotlib.pyplot as plt import pandas as pd

# Air quality dataset

data = {

'Location 1': [20, 30, 35, 45, 50, 60, 70, 80, 90, 100],

 'Location 2': [15, 25, 30, 40, 50, 65, 75, 85, 95, 105],

'Location 3': [25, 35, 40, 50, 60, 75, 85, 95, 105, 115]

}


# Create a DataFrame from the dataset

df = pd.DataFrame(data)


# Create a box plot

plt.figure(figsize=(8, 6))

plt.boxplot(df.values, labels=df.columns, sym='b+')

plt.title('Air Quality Box Plot')
```

```
plt.xlabel('Monitoring Locations')
plt.ylabel('Pollutant Concentration')
plt.grid(True)

# Display the box plot
plt.show()
```

OUTPUT:



BUBBLE CHART:

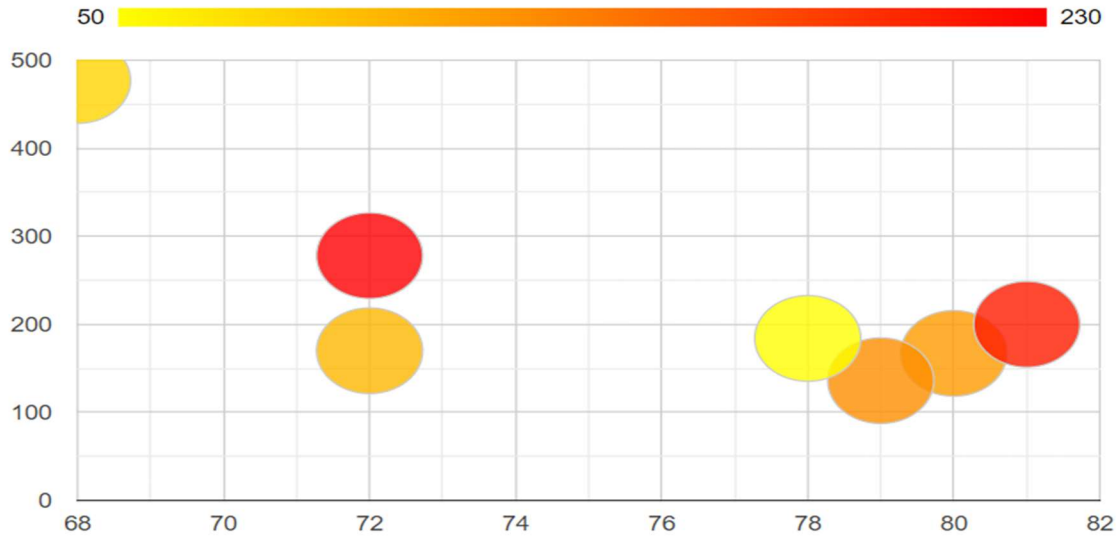
A bubble chart is a type of data visualization that extends the concept of a scatter plot. It is used to display three-dimensional data in a two-dimensional graph. In a bubble chart, data points are represented as bubbles (or circles) on a graph, and the size of each bubble is used to convey the value of a third variable, in addition to the standard x and y coordinates.

PROGRAM:

```
import pandas as pd import matplotlib.pyplot as plt

# Air quality dataset

data = { 'Location': ['Chennai', 'Coimbatore', 'Madurai', 'Trichy','Salem'],

'PM2.5': [15, 22, 18, 20, 17],

'NO2': [10, 12, 11, 13, 9],

'Population': [4646732, 1642658, 1479978, 1065138, 873363] }

df = pd.DataFrame(data)


# Parameters for bubble chart

x = 'PM2.5'  # X-axis variable (e.g., PM2.5 concentration)

y = 'NO2'  # Y-axis variable (e.g., NO2 concentration)

s = 'Population'  # Size of the bubbles (e.g., population)

location = 'Location' # Location labels
```

```
# Create the bubble chart
plt.figure(figsize=(10, 6))
plt.scatter(df[x], df[y], s=df[s] / 10000, c='b', alpha=0.6, edgecolors='w', linewidth=1)

# Adding labels to data points
for i, txt in enumerate(df[location]):
plt.annotate(txt, (df[x][i], df[y][i]), fontsize=12, ha='center')

# Adding labels and title
 plt.xlabel(x)
plt.ylabel(y)
plt.title(f'Bubble Chart for Air Quality Analysis')

# Show the chart
plt.show()
```

OUTPUT:



CLUSTERED BAR:

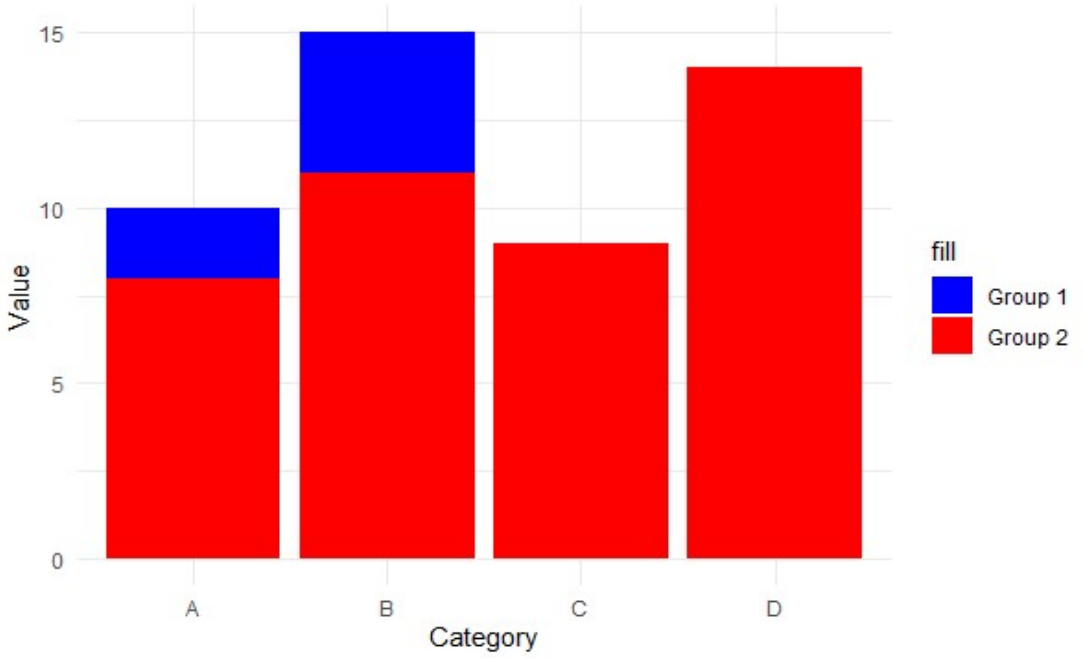
A clustered bar chart is a type of data visualization that displays data using rectangular bars grouped in clusters. In this chart, each cluster represents a category or a set of data points, and within each cluster, multiple bars are used to show subcategories or different data series. Clustered bar charts are commonly used to compare values across categories and subcategories, allowing for easy visual comparison of data within and between clusters.

PROGRAM:

```
import pandas as pd
import matplotlib.pyplot as plt

# Air quality dataset
data = {
    'Station': ['Station A', 'Station B', 'Station C', 'Station D'],
    'NO2': [25, 20, 30, 15],
    'SO2': [10, 15, 8, 12],
    'PM2.5': [40, 35, 45, 30],
}

# Create a DataFrame from the sample data
df = pd.DataFrame(data)

# Define the pollutants and their colors for the chart
pollutants = ['NO2', 'SO2', 'PM2.5']
colors = ['#ff5733', '#33ff57', '#5733ff']

# Set the width of the bars
bar_width = 0.2

# Set the positions for the x-axis ticks
x = range(len(df['Station']))

# Create a clustered bar chart
fig, ax = plt.subplots(figsize=(10, 6))

for i, pollutant in enumerate(pollutants):
    ax.bar(
        [pos + i * bar_width for pos in x],
```

```
        df[pollutant],
        width=bar_width,
        label=pollutant,
        color=colors[i],
    )

# Set the x-axis labels
ax.set_xticks([pos + (len(pollutants) - 1) * bar_width / 2 for pos in x])
ax.set_xticklabels(df['Station'])

# Set labels and title
ax.set_xlabel('Monitoring Stations')
ax.set_ylabel('Concentration (µg/m³)')
ax.set_title('Air Quality Analysis for Different Pollutants')

# Add a legend
ax.legend()

# Show the plot
plt.tight_layout()
plt.show()
```

OUTPUT:



CLUSTERED COLUMN:

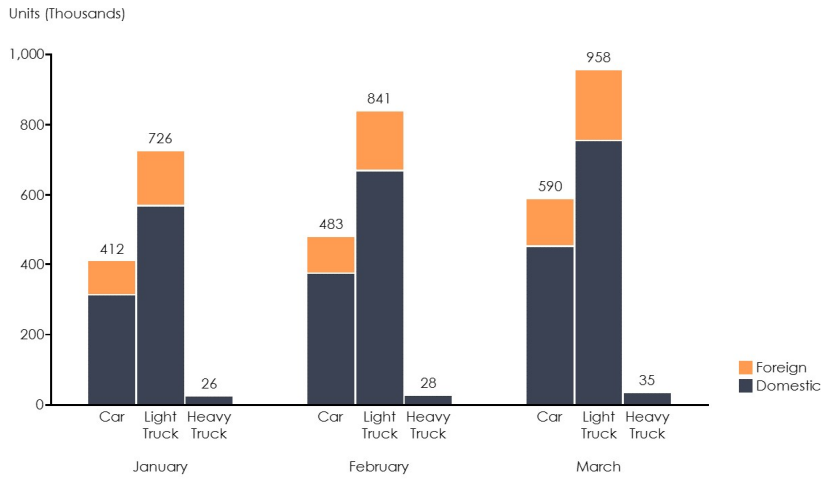
A clustered column chart is a type of data visualization that displays data using vertical rectangular columns grouped in clusters. In this chart, each cluster represents a category or a set of data points, and within each cluster, multiple columns are used to show subcategories or different data series. Clustered column charts are commonly used to compare values across categories and subcategories, allowing for easy visual comparison of data within and between clusters.

PROGRAM:

```
import matplotlib.pyplot as plt
import numpy as np

# Air quality dataset
stations = ['Station A', 'Station B', 'Station C']
pollutants = ['PM2.5', 'NO2', 'SO2']
concentrations = np.array([[10, 15, 12],
                           [8, 11, 9],
                           [20, 18, 22]])

# Number of clusters and the width of each cluster
n_clusters = len(stations)
cluster_width = 0.2

# Create a figure and axis
fig, ax = plt.subplots()

# Generate x values for the clusters
x = np.arange(len(pollutants))

# Create clustered column chart
for i in range(n_clusters):
    ax.bar(x + i * cluster_width, concentrations[i], width=cluster_width,
label=stations[i])

# Set x-axis labels
ax.set_xticks(x + (cluster_width * (n_clusters - 1)) / 2)
ax.set_xticklabels(pollutants)
ax.set_xlabel('Pollutants')

# Set y-axis label
```

```
ax.set_ylabel('Concentration (µg/m³)')

# Add a legend
ax.legend()

# Set the title and grid
plt.title('Air Quality Analysis by Pollutant and Monitoring Station')
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Show the chart
plt.show()
```

OUTPUT:



CONCLUSION

- In conclusion, the visualization and analysis of air quality data in Tamil Nadu have provided valuable insights into the state of our environment and its implications for public health and urban development.

- The visualization and analysis of air quality data are pivotal in addressing the growing concerns about air pollution in our region. By employing data-driven insights, we can work collectively to create a cleaner, healthier, and more sustainable environment for current and future generations. The significance of this work cannot be overstated, as it directly impacts the well-being and quality of life for all residents of Tamil Nadu.

- In the era of increasing environmental awareness and the growing importance of public health, the analysis and visualization of air quality data have become indispensable tools. These practices empower governments, organizations, and individuals to take meaningful actions toward cleaner, healthier air for all. As we move forward, it is essential to continue refining our analytical techniques, enhancing data quality, and leveraging technology to better understand and address air quality challenges worldwide.