PROJECT-2

AIR QUALITY ASSESSMENT TN

INTRODUCTION

Air quality refers to the condition or cleanliness of the air in a specific location, typically measured in terms of the presence and concentration of various air pollutants.

The quality of the air we breathe is of paramount importance, as it directly impacts human health, the environment, and overall well-being. This document aims to provide a comprehensive description of air quality, including the factors influencing it, key pollutants, and the methods used for measuring and assessing air quality.

Air quality depends on a combination of natural and human-made factors. These factors include emissions from various sources, such as industrial processes and energy production, which significantly affect air quality. Additionally, weather patterns, climate change, and chemical reactions in the atmosphere can also lead to the formation of secondary pollutants, such as harmful gases like ozone, which have a substantial impact on air quality.

ABSTRACT

Air quality data analysis provides a powerful and unique method for extracting valuable insights from the extensive information collected through air quality monitoring stations and sensors in various regions. This analysis allows for the identification of patterns in air pollution data and informs decision-making processes related to pollution control measures.

In this document, we will explore the data analysis process for air quality and the associated methods.

DATA SOURCES

The data required for this analysis consists of historical air quality information for a specific region or over a set of years. These data sources include data from air quality monitoring stations and satellite remote sensing equipment for visualization.

ANALYSIS OF THE PROBLEM - STEPS :

1. To initiate the data analysis, air quality data for the Tamil Nadu region was collected. The dataset underwent preprocessing, which involved handling missing data, removing outliers, and ensuring data consistency.

2. Subsequently, the air pollution dataset was analyzed through various visualization techniques, such as bar charts, pie charts, line plots, and bar plots. This aided in understanding the relationships between air quality parameters and factors like weather and climate change.

3. Spatial distribution of air quality was achieved by identifying pollution hotspots, predicting air quality levels, and pinpointing pollution sources. This information facilitated the optimization of pollution control strategies, incorporating methods like clustering and neural networks.

4. During the decision-making process using the dataset, trend analysis was a crucial consideration. This involved identifying long-term trends in air quality data to assess whether air quality was improving or deteriorating over time. Additionally, forecasting was employed to predict future patterns based on historical data and current information, aiding in more effective decision-making for areas affected by pollution.

5. This comprehensive data analysis process contributes to a deeper understanding of air quality and supports informed decisions for pollution control and public health protection.

```
# Importing the necessary python libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Importing the csv file
df = pd.read_csv('Air_Quality.csv')
df.head()
```

Output:

	Stn Code	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO2	NO2	RSPM/PM10	SPM
0	38	5/1/2010	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	9.60	17.166667	73.333333	149.666667
1	38	7/1/2010	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.15	20.283333	61.333333	150.333333
2	38	12/1/2010	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.45	20.516667	75.000000	114.666667
3	38	1/19/2010	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	10.75	18.183333	120.000000	197.666667
4	38	1/21/2010	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	9.78	17.320000	96.500000	216.000000

```
df.shape
```

Output:

```
(12351, 10)
```

```
df.info()
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12351 entries, 0 to 12350
Data columns (total 10 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Stn Code                        12351 non-null  int64
 1   Sampling Date                   12351 non-null  object
 2   City/Town/Village/Area          12351 non-null  object
 3   Location of Monitoring Station  12351 non-null  object
 4   Agency                          12351 non-null  object
 5   Type of Location                12351 non-null  object
 6   SO2                             12149 non-null  float64
 7   NO2                             12153 non-null  float64
 8   RSPM/PM10                       12303 non-null  float64
 9   SPM                             1902 non-null   float64
dtypes: float64(4), int64(1), object(5)
memory usage: 965.0+ KB
```

```
# Removing duplicate values
df.drop_duplicates()
```

Output:

```
1
```

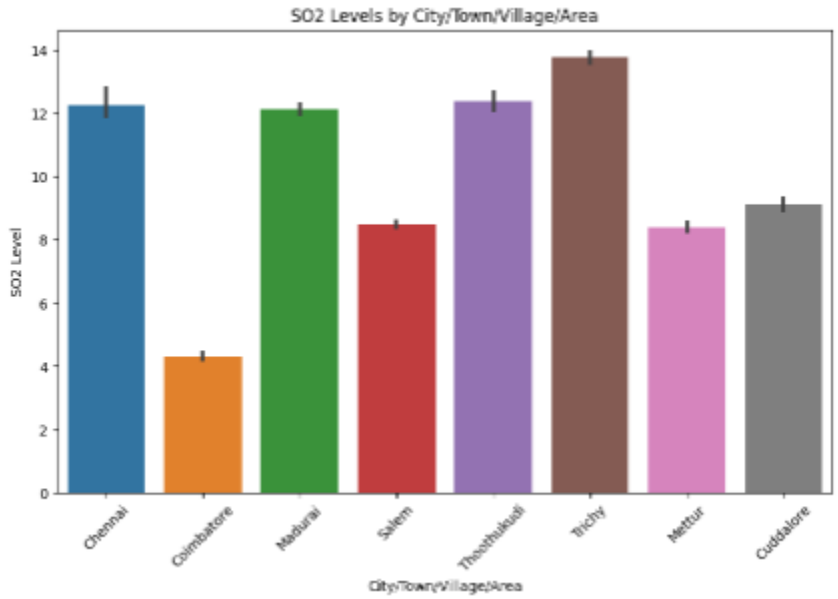
```
# Check for null values
df.isna().sum()
#pd.isnull(df).sum()
```

Output:

```
Stn Code                          0
Sampling Date                     0
City/Town/Village/Area            0
Location of Monitoring Station    0
Agency                            0
SO2                             202
NO2                             198
RSPM/PM10                        48
SPM                           10449
dtype: int64
```

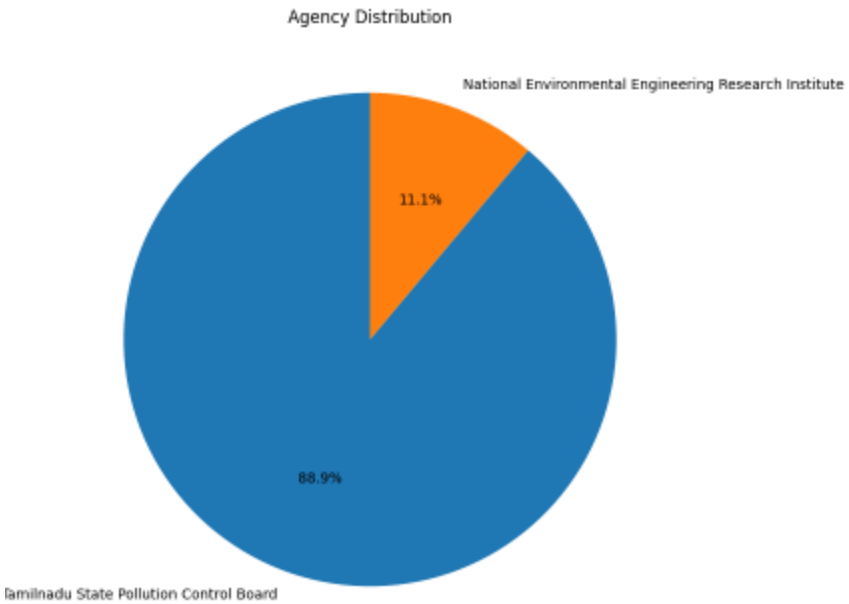
```
# Create the line chart
dat = pd.DataFrame(df)
plt.figure(figsize=(10, 6))
sns.barplot(x='City/Town/Village/Area', y='SO2', data=dat)
plt.title('SO2 Levels by City/Town/Village/Area')
plt.xticks(rotation=45)
plt.xlabel('City/Town/Village/Area')
plt.ylabel('SO2 Level')
plt.show()
```

Output:




```
# Pie chart for 'Agency' column
agency_counts = df['Agency'].value_counts()
plt.figure(figsize=(8, 8))
plt.pie(agency_counts, labels=agency_counts.index, autopct='%1.1f%%',
startangle=90)
plt.title('Agency Distribution')
plt.show()
```

Output:



```
# Scatter plot for 'SO2' vs 'NO2'
plt.figure(figsize=(10, 6))
plt.scatter(df['SO2'], df['NO2'], c='blue', alpha=0.5)
plt.title('SO2 vs NO2 Scatter Plot')
plt.xlabel('SO2')
plt.ylabel('NO2')
plt.grid(True)  # Add gridlines (optional)
plt.show()
```

Output: