SQL + EXCEL POLLUTION PROJECT

"Shared Responsibility for a Healthy Environment"

INTRODUCTION

This document outlines key questions and insights derived from the analysis of a comprehensive pollution dataset. The dataset contains vital environmental metrics, including PM2.5, PM10, nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), temperature, and humidity. These metrics are analyzed to uncover patterns, correlations, and potential risk factors affecting air quality in various regions.

INTRODUCTION

The analysis is categorized into three sections—basic, intermediate, and advanced—each addressing progressively more complex questions. Basic questions focus on fundamental metrics such as pollutant averages and air quality classifications. Intermediate questions delve into relationships between environmental factors and pollutants, while advanced questions explore multifaceted interactions and predictive modeling opportunities.

INTRODUCTION

This structured approach provides actionable insights into air quality dynamics, helping stakeholders understand critical environmental trends and their impact on public health.

"ROLE OF EXCEL IN DATA PREPARATION AND ANALYSIS"

Microsoft Excel played a crucial role in this project by facilitating efficient data cleaning, organization, and initial analysis. It was used to generate a unique RecordID column, ensuring every row had a primary key for easy identification. Null values were handled effectively using Excel's "Go To Special" and "Blank" features, while duplicate records were checked to maintain data integrity. The dataset was systematically organized, enabling easy filtering and sorting based on parameters like pollutant levels, industrial proximity, and air quality categories.

"ROLE OF EXCEL IN DATA PREPARATION AND ANALYSIS"

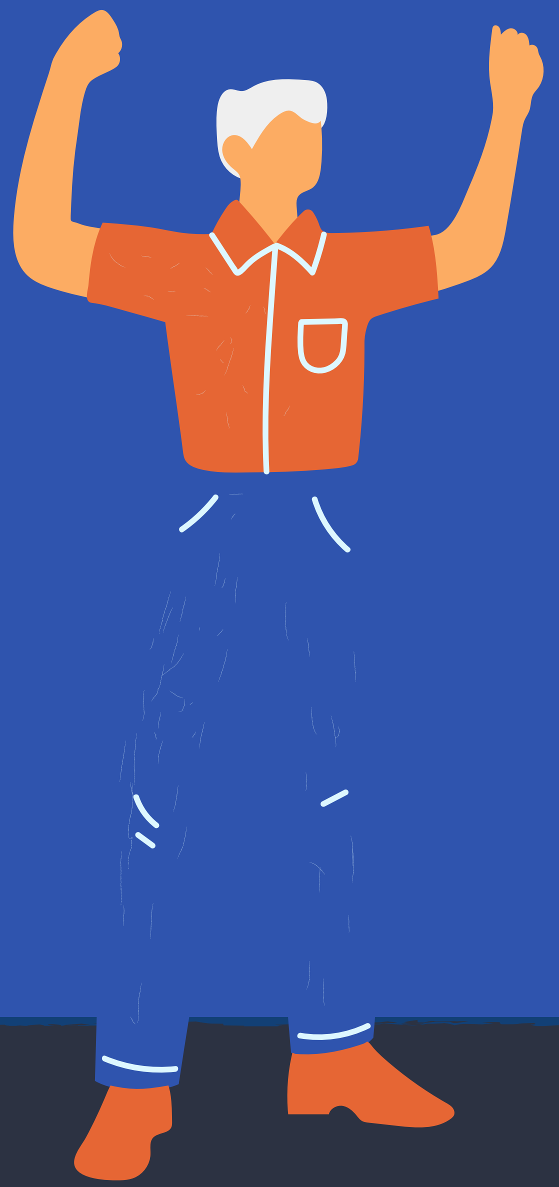
Excel’s built-in functions, such as AVERAGE, MIN, and MAX, were employed to calculate basic statistics. Additionally, Excel served as a bridge to SQL analysis by exporting the cleaned and formatted dataset as a CSV file, ensuring consistency and accuracy. By leveraging Excel's user-friendly interface and powerful features, the project achieved a seamless data preparation process, paving the way for advanced SQL-based analysis.

SQL + EXCEL UPDATED POLLUTION PROJECT

```
--Sql updated pollution Project
CREATE DATABASE updated_pollution_project;
```

CREATE TABLE

SICK OF POLLUTION



```sql
--Create Table
CREATE TABLE pollution_data (
    RecordID INT PRIMARY KEY,
    Temperature FLOAT,
    Humidity FLOAT,
    PM25_Concentration FLOAT,
    PM10_Concentration FLOAT,
    Nitrogen_Dioxide_Levels FLOAT,
    Sulfur_Dioxide_Levels FLOAT,
    Carbon_Monoxide_Levels FLOAT,
    Proximity_to_Industrial_Areas FLOAT,
    Population_Density INT,
    Air_Quality VARCHAR(9)  );
```

TABLE

	recordid [PK] integer	temperature double precision	humidity double precision	pm25_concentration double precision	pm10_concentration double precision	nitrogen_dioxide_levels double precision	sulfur_dioxide_levels double precision	carbon_monoxide_levels double precision	proximity_to_industrial_areas double precision	population_density integer	air_quality character varying (9)
1	1	29.8	59.1	5.2	17.9	18.9	9.2	1.72	6.3	319	Moderate
2	2	28.3	75.6	2.3	12.2	30.8	9.7	1.64	6	611	Moderate
3	3	23.1	74.7	26.7	33.8	24.4	12.6	1.63	5.2	619	Moderate
4	4	27.1	39.1	6.1	6.3	13.5	5.3	1.15	11.1	551	Good
5	5	26.5	70.7	6.9	16	21.9	5.6	1.01	12.7	303	Good
6	6	39.4	96.6	14.6	35.5	42.9	17.9	1.82	3.1	674	Hazardous
7	7	41.7	82.5	1.7	15.8	31.1	12.7	1.8	4.6	735	Poor
8	8	31	59.6	5	16.8	24.2	13.6	1.38	6.3	443	Moderate
9	9	29.4	93.8	10.3	22.7	45.1	11.8	2.03	5.4	486	Poor
10	10	33.2	80.5	11.1	24.4	32	15.3	1.69	4.9	535	Poor
11	11	26.3	65.7	1.3	5.5	18.3	5.9	0.85	13	529	Good
12	12	32.5	51.2	1.6	10.5	21.6	19.3	1.53	5.9	519	Moderate
13	13	20	53.3	3.7	12.9	26.1	6.6	1.09	10.2	538	Good
14	14	28.6	53.7	28.9	34	23.2	4.5	1.02	11	508	Good
15	15	22.3	80.5	4.5	12	17.2	6.3	1.18	10.4	232	Good
16	16	32	78.9	22.4	29.9	27.5	11.8	1.48	7.9	444	Moderate
17	17	22.9	75.4	4.5	10.4	18.4	3.7	0.96	14.4	359	Good
18	18	37.6	81.2	28.1	56.6	46.7	13.7	1.85	4.1	560	Poor
19	19	34.7	59.3	9	15.7	28.5	7.1	1.52	6.1	437	Moderate
20	20	37.8	97.2	0.6	24.6	37.1	11.7	1.13	7.7	803	Poor
21	21	27.6	44.1	3.5	14.4	30.7	9.4	0.97	8	338	Moderate
22	22	27.6	77.5	3.8	10.9	9.1	1.7	1.04	14.4	520	Good
23	23	25.6	58.3	0.4	-0.2	25.3	4.5	0.98	10	536	Good
24	24	24.6	48.4	8.3	15.4	23.3	4.6	1.03	11.2	461	Good
25	25	26.2	49.9	0.5	2	22.9	2.3	0.94	17.9	581	Good
26	26	22.6	62.1	16.4	21.3	17.6	4.8	0.93	11.5	324	Good
27	27	30	69.2	2	3.9	15.3	5.2	1.09	10.1	297	Good
28	28	37.6	72.2	10.1	16.6	32.2	6.3	1.85	3.5	637	Poor
29	29	29.6	65.6	3.4	9.3	20.6	2.9	1.03	10.2	452	Good
30	30	52.5	83.9	67.5	91.2	31.6	34.6	2.31	2.9	785	Hazardous
31	31	44.7	97.9	34.6	57.3	28.1	22.6	1.78	3.5	562	Hazardous

DATA CLEANING IN EXCEL

--Data Cleaning in Excel

--Generated a Primary Key Column:
--Created a new column named RecordID to serve as the primary key, as the dataset originally lacked a primary key.

--Removed Unnecessary Rows:

--Deleted 3 rows that were deemed unnecessary.

--Checked for Duplicates:
--Performed a check for duplicate rows, but no duplicates were found in the dataset.

--Deleted Null Values:

--Identified and removed rows containing null values using Excel's "Go To Special" feature:
--Selected all blank cells.
--Right-clicked and deleted the corresponding rows.

DATA CLEANING IN EXCEL

--"Renaming Columns for Clarity and Compatibility"
--The column names in the dataset were updated for better understanding.
--NO2 was changed to Nitrogen_Dioxide_Levels to represent nitrogen dioxide levels clearly.
--Similarly, SO2 became Sulfur_Dioxide_Levels, and CO was renamed Carbon_Monoxide_Levels to avoid confusion with carbon dioxide.
--The particulate matter columns were updated as well: PM10 is now PM10_Concentration, and PM2.5 is PM25_Concentration,
--ensuring clarity and compatibility with database systems.
--These changes make the dataset more intuitive for analysis and reporting.

DATA CLEANING IN EXCEL

-- Note: The column "PM25_CONCENTRATION" in the dataset refers to "PM2.5 Concentration."
-- The naming avoids using a period (".") for compatibility with database systems.

--Here’s a Brief Explanation.
--The column "PM25_CONCENTRATION" in the dataset represents "PM2.5 Concentration," which refers
--to particulate matter with a diameter of 2.5 micrometers or smaller. These fine particles are a critical
--pollutant affecting air quality and public health, measured in micrograms per cubic meter (µg/m³).
--Due to technical constraints in database systems, where special characters like periods (".") are not
--supported in column names, "PM2.5" was renamed to "PM25" for compatibility. While the naming
--convention has been modified, the data still reflects the same information about PM2.5 levels.
--This adjustment ensures smooth querying in SQL environments without altering the data's meaning
or relevance.

CHECK NULL VALUES

```
--check null values
SELECT * FROM pollution_data
WHERE recordid IS NULL OR
temperature IS NULL OR
humidity IS NULL OR
pm25_CONCENTRATION IS NULL OR
PM10_CONCENTRATION IS NULL OR
NITROGEN_DIOXIDE_LEVELS IS NULL OR
SULFUR_DIOXIDE_LEVELS IS NULL OR
CARBON_MONOXIDE_LEVELS IS NULL OR
PROXIMITY_TO_INDUSTRIAL_AREAS IS NULL OR
POPULATION_DENSITY IS NULL OR
AIR_QUALITY IS NULL ;
```

DATA EXPLORATION

```sql
-- Data Exploration

-- How many Data we have?
 SELECT COUNT(*) AS total_record FROM POLLUTION_DATA;
```

total_record bigint
4996

DATA EXPLORATION

```sql
--What is the distribution of temperature and humidity in the dataset?

 SELECT
     ROUND(AVG(Temperature)::numeric ,2) AS Avg_Temperature,
     MIN(Temperature) AS Min_Temperature,
     MAX(Temperature) AS Max_Temperature,
     ROUND(AVG(Humidity)::numeric ,2) AS Avg_Humidity,
     MIN(Humidity) AS Min_Humidity,
     MAX(Humidity) AS Max_Humidity
 FROM Pollution_Data;
```

avg_temperature numeric	min_temperature double precision	max_temperature double precision	avg_humidity numeric	min_humidity double precision	max_humidity double precision
30.03	13.4	58.6	70.06	36	128.1

DATA EXPLORATION

```
--What are the average, minimum, and maximum values of PM2.5, PM10, NO₂, SO₂, and CO?

 SELECT
     ROUND(AVG(PM25_Concentration)::numeric ,2) AS Avg_PM25,
     MIN(PM25_Concentration) AS Min_PM25,
     MAX(PM25_Concentration) AS Max_PM25,
     ROUND(AVG(PM10_Concentration)::numeric ,2) AS Avg_PM10,
     MIN(PM10_Concentration) AS Min_PM10,
     MAX(PM10_Concentration) AS Max_PM10,
     ROUND(AVG(Nitrogen_dioxide_levels)::numeric ,2) AS Avg_NO2,
     MIN(Nitrogen_dioxide_levels) AS Min_NO2,
     MAX(Nitrogen_dioxide_levels) AS Max_NO2,
     ROUND(AVG(Sulfur_dioxide_levels)::numeric ,2) AS Avg_SO2,
     MIN(Sulfur_dioxide_levels) AS Min_SO2,
     MAX(Sulfur_dioxide_levels) AS Max_SO2,
     ROUND(AVG(Carbon_monoxide_levels)::numeric ,2) AS Avg_CO,
     MIN(Carbon_monoxide_levels) AS Min_CO,
     MAX(Carbon_monoxide_levels) AS Max_CO
 FROM Pollution_Data;
```

DATA EXPLORATION

avg_pm25 numeric	min_pm25 double precision	max_pm25 double precision	avg_pm10 numeric	min_pm10 double precision	max_pm10 double precision	avg_no2 numeric	min_no2 double precision	max_no2 double precision	avg_so2 numeric	min_so2 double precision	max_so2 double precision	avg_co numeric	min_co double precision	max_co double precision
20.15	0	295	30.23	-0.2	315.8	26.42	7.4	64.9	10.02	-6.2	44.9	1.50	0.65	3.72

DATA ANALYSIS & BUSINESS KEY PROBLEMS & ANSWERS

```
-- Data Analysis & Business Key Problems & Answers_

-- ---------------------------------------------------------------------------------------------------------------------
-- | Question No. | Question                                                                                           |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.1          | How many records are present in the dataset?                                                       |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.2          | What is the average PM2.5 and PM10 concentration recorded in the dataset?                          |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.3          | What is the average temperature and humidity recorded across all entries in the dataset?           |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.4          | How many areas are classified under each air quality category such as Good or Moderate?            |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.5          | How many areas are in close proximity to industrial zones based on recorded distances?             |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Intermediate Questions                                                                                            |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.6          | Which air quality category reports the highest average PM2.5 concentration in the given dataset?   |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.7          | What percentage of areas with high population density (> 500) have moderate or poor air quality?   |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.8          | How does air quality differ in areas with high humidity (> 70) compared to low humidity (< 30)?    |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.9          | What is the relationship between industrial zone proximity and sulfur dioxide levels in the dataset? |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.10         | How does population density correlate with carbon monoxide levels in different regions?            |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Advanced Questions                                                                                                |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.11         | Which pollutants (PM2.5, PM10, NO₂, SO₂, CO) frequently exceed safe thresholds when measured       |
-- |              | together?                                                                                          |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.12         | How does the combination of high PM2.5 and PM10 concentrations affect the air quality classification?|
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.13         | What impact does industrial zone proximity have on air quality when accounting for population density?|
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.14         | Can air quality classification (e.g., Good, Moderate) be predicted based on pollutant and          |
-- |              | environmental data?                                                                                |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.15         | Which demographic factors (e.g., high population density, proximity to industry) indicate a higher |
-- |              | risk of poor air quality?                                                                          |
-- |--------------|----------------------------------------------------------------------------------------------------|
-- | Q.16         | What percentage of records show "Good" air quality in areas with proximity to industrial zones?    |
-- |--------------|----------------------------------------------------------------------------------------------------|
```

Q.1 HOW MANY RECORDS ARE PRESENT IN THE DATASET?

```sql
--Q.1 How many records are present in the dataset?

SELECT COUNT(*) AS Total_Records
FROM pollution_data;
```

total_records bigint
4996

Q.2 WHAT IS THE AVERAGE PM2.5 AND PM10 CONCENTRATION RECORDED IN THE DATASET?

```sql
--Q.2 What is the average PM2.5 and PM10 concentration recorded in the dataset?

SELECT ROUND(AVG(pm25_concentration)::numeric, 2) AS avg_pm2_5,
 ROUND(AVG(pm10_concentration)::numeric, 2) AS avg_pm10
FROM pollution_data;
```

avg_pm2_5 numeric	avg_pm10 numeric
20.15	30.23

Q.3 WHAT IS THE AVERAGE TEMPERATURE AND HUMIDITY RECORDED ACROSS ALL ENTRIES IN THE DATASET?

```
--Q.3 What is the average temperature and humidity recorded across all entries in the dataset?

 SELECT ROUND(AVG(temperature)::numeric, 2) AS avg_temperature,
  ROUND(AVG(humidity)::numeric, 2) AS avg_humidity
 FROM pollution_data;
```

avg_temperature numeric	avg_humidity numeric
30.03	70.06

Q.4 HOW MANY AREAS ARE CLASSIFIED UNDER EACH AIR QUALITY CATEGORY SUCH AS GOOD OR MODERATE?

```
--Q.4 How many areas are classified under each air quality category such as Good or Moderate?

SELECT
    Air_Quality,
    COUNT(*) AS area_count
FROM Pollution_Data
GROUP BY Air_Quality
ORDER BY area_count DESC;
```

air_quality character varying (9)	area_count bigint
Good	1997
Moderate	1499
Poor	1000
Hazardous	500

Q.5 HOW MANY AREAS ARE IN CLOSE PROXIMITY TO INDUSTRIAL ZONES BASED ON RECORDED DISTANCES?

```sql
--Q.5 How many areas are in close proximity to industrial zones based on recorded distances?

SELECT COUNT(*) AS High_Proximity_Areas FROM pollution_data
WHERE proximity_to_industrial_areas <=10;
```

high_proximity_areas bigint
2870

Q.6 WHICH AIR QUALITY CATEGORY REPORTS THE HIGHEST AVERAGE PM2.5 CONCENTRATION IN THE GIVEN DATASET?

```sql
--Q.6 Which air quality category reports the highest average PM2.5 concentration in the given dataset?

SELECT
    Air_Quality,
    round(AVG(PM25_Concentration)::numeric ,2) AS Avg_PM2_5
FROM Pollution_Data
GROUP BY Air_Quality
ORDER BY Avg_PM2_5 DESC
LIMIT 1;
```

air_quality character varying (9)	avg_pm2_5 numeric
Hazardous	41.92

Q.7 WHAT PERCENTAGE OF AREAS WITH HIGH POPULATION DENSITY (> 500) HAVE MODERATE OR POOR AIR QUALITY?

```sql
--Q.7 What percentage of areas with high population density (> 500) have moderate or poor air quality?

SELECT
    COUNT(*) * 100.0 / (SELECT COUNT(*) FROM pollution_data
    WHERE Population_Density > 500) AS Percentage
    FROM Pollution_Data
    WHERE Population_Density > 500 AND Air_Quality IN ('Moderate', 'Poor');
```

percentage numeric
60.4612850082372323

Q.8 HOW DOES AIR QUALITY DIFFER IN AREAS WITH HIGH HUMIDITY (> 70) COMPARED TO LOW HUMIDITY (< 30)?

```sql
--Q.8 How does air quality differ in areas with high humidity (> 70) compared to low humidity (< 30)?

SELECT
    CASE
        WHEN Humidity > 70 THEN 'High Humidity'
        WHEN Humidity < 30 THEN 'Low Humidity'
    END AS Humidity_Category,
     Air_Quality,
    COUNT(*) AS Count
FROM Pollution_Data
GROUP BY Humidity_Category, Air_Quality;
```

humidity_category text	air_quality character varying (9)	count bigint
High Humidity	Moderate	777
[null]	Poor	249
High Humidity	Hazardous	462
High Humidity	Poor	751
[null]	Hazardous	38
[null]	Good	1513
High Humidity	Good	484
[null]	Moderate	722

Q.9 WHAT IS THE RELATIONSHIP BETWEEN INDUSTRIAL ZONE PROXIMITY AND SULFUR DIOXIDE LEVELS IN THE DATASET?

```sql
--Q.9 What is the relationship between industrial zone proximity and sulfur dioxide levels in the dataset?

SELECT
    Proximity_to_Industrial_Areas,
    Round(AVG(Sulfur_dioxide_levels)::numeric ,2) AS Avg_SO2
FROM Pollution_Data
GROUP BY Proximity_to_Industrial_Areas
ORDER BY Proximity_to_Industrial_Areas;
```

	proximity_to_industrial_areas double precision	avg_so2 numeric
1	2.5	19.02
2	2.6	16.81
3	2.7	19.40
4	2.8	20.61
5	2.9	20.24
6	3	20.21
7	3.1	18.62
8	3.2	19.57
9	3.3	16.67
10	3.4	18.48
11	3.5	17.49

Q.10 HOW DOES POPULATION DENSITY CORRELATE WITH CARBON MONOXIDE LEVELS IN DIFFERENT REGIONS?

```sql
--Q.10 How does population density correlate with carbon monoxide levels in different regions?

SELECT
    Population_Density,
    Round(AVG(Carbon_monoxide_levels)::numeric ,2) AS Avg_CO
FROM Pollution_Data
GROUP BY Population_Density
ORDER BY Population_Density;
```

	population_density integer	avg_co numeric
1	188	1.00
2	189	0.96
3	191	0.89
4	193	1.01
5	194	1.08
6	196	1.07
7	197	0.89
8	198	1.06
9	199	0.97
10	200	0.92
11	201	0.98
12	202	1.00
13	203	0.96
14	204	1.04
15	205	1.01
16	206	1.01
17	207	1.02
18	208	1.00

Q.11 WHICH POLLUTANTS (PM2.5, PM10, NO_2, SO_2, CO) FREQUENTLY EXCEED SAFE THRESHOLDS WHEN MEASURED TOGETHER?

```sql
--Q.11 Which pollutants (PM2.5, PM10, NO₂, SO₂, CO) frequently exceed safe thresholds when measured together?

SELECT
    COUNT(*) AS Frequency
FROM Pollution_data
WHERE PM25_Concentration > 25
      AND PM10_Concentration > 50
      AND Nitrogen_dioxide_levels > 40
      AND Sulfur_dioxide_levels > 20
      AND Carbon_monoxide_levels > 1.0;
```

frequency bigint
80

Q.11 WHICH POLLUTANTS (PM2.5, PM10, NO_2, SO_2, CO) FREQUENTLY EXCEED SAFE THRESHOLDS WHEN MEASURED TOGETHER?

```
--Note:
--The thresholds used in the SQL query are based on air quality standards or guidelines set by organizations
--such as the World Health Organization (WHO) or local regulatory bodies. These thresholds are typically defined
--as the maximum safe limits for specific pollutants to minimize harm to human health. Here’s why these values might be chosen:
--Why Use These Values?
--These thresholds represent safe exposure limits. By checking which pollutants exceed these limits together,
--the query helps identify combinations of harmful air quality factors that occur simultaneously.
--If you want to change these limits (based on local standards), you can adjust the values in the query accordingly. For example:
--If your region's limit for PM2.5 is 35 µg/m³ instead of 25, you would update the query condition to PM2.5 Concentration > 35.
```

Q.12 HOW DOES THE COMBINATION OF HIGH PM2.5 AND PM10 CONCENTRATIONS AFFECT THE AIR QUALITY CLASSIFICATION?

```sql
--Q.12 How does the combination of high PM2.5 and PM10 concentrations affect the air quality classification?

SELECT
    Air_Quality,
    COUNT(*) AS Count
FROM Pollution_Data
WHERE PM25_Concentration > 25 AND PM10_Concentration > 50
GROUP BY Air_Quality;
```

air_quality character varying (9)	count bigint
Good	17
Moderate	229
Poor	303
Hazardous	245

Q.13 WHAT IMPACT DOES INDUSTRIAL ZONE PROXIMITY HAVE ON AIR QUALITY WHEN ACCOUNTING FOR POPULATION DENSITY?

```sql
--Q.13 What impact does industrial zone proximity have on air quality when accounting for population density?

 SELECT
     Proximity_to_Industrial_Areas,
     Population_Density,
     AVG(PM25_Concentration) AS Avg_PM2_5,
     AVG(PM10_Concentration) AS Avg_PM10,
     Air_Quality
 FROM Pollution_data
 GROUP BY Proximity_to_Industrial_Areas, Population_Density, Air_Quality;
```

Q.13 WHAT IMPACT DOES INDUSTRIAL ZONE PROXIMITY HAVE ON AIR QUALITY WHEN ACCOUNTING FOR POPULATION DENSITY?

proximity_to_industrial_areas double precision	population_density integer	avg_pm2_5 double precision	avg_pm10 double precision	air_quality character varying (9)
6.6	553	3.8	7.1	Moderate
10.1	365	26.9	33.8	Good
5.1	524	7	9.5	Moderate
7.2	418	46.3	58.9	Poor
2.8	631	0.2	26.8	Hazardous
8.8	393	7.7	12.4	Moderate
12	317	19.2	26.3	Good
10.5	267	2	6.3	Good
3	708	99.2	115.1	Hazardous
4.8	420	6.2	20.8	Poor
10.1	265	6.5	11.1	Good
18	559	6.1	12.5	Good
5.8	591	29.3	40.2	Moderate
6.7	707	95.7	115.6	Hazardous
6	268	25.8	32.4	Moderate
8.3	813	35.8	67	Hazardous
7.3	440	10.3	15.5	Moderate
15.9	312	3.7	10.1	Good
4.7	688	11.3	27.2	Hazardous
7.2	296	34.4	39.6	Moderate
10.5	441	9.9	14.9	Good

Q.14 CAN AIR QUALITY CLASSIFICATION (E.G., GOOD, MODERATE) BE PREDICTED BASED ON POLLUTANT AND ENVIRONMENTAL DATA?

```sql
--Q.14 Can air quality classification (e.g., Good, Moderate) be predicted based on pollutant and environmental data?

 SELECT
     Air_Quality,
     AVG(PM25_Concentration) AS Avg_PM2_5,
     AVG(PM10_Concentration) AS Avg_PM10,
     AVG(Temperature) AS Avg_Temperature,
     AVG(Humidity) AS Avg_Humidity
 FROM Pollution_Data
 GROUP BY Air_Quality;
```

air_quality character varying (9)	avg_pm2_5 double precision	avg_pm10 double precision	avg_temperature double precision	avg_humidity double precision
Good	9.913870806209296	14.992538808212332	24.952178267401134	60.021832749123696
Moderate	20.474783188792525	30.61714476317543	30.135823882588376	70.22121414276198
Poor	29.235799999999948	44.4521	34.86510000000001	80.18289999999999
Hazardous	41.9208	61.50799999999993	40.34779999999998	89.47299999999998

Q.15 WHICH DEMOGRAPHIC FACTORS (E.G., HIGH POPULATION DENSITY, PROXIMITY TO INDUSTRY) INDICATE A HIGHER RISK OF POOR AIR QUALITY?

```sql
--Q.15  Which demographic factors (e.g., high population density, proximity to industry) indicate a higher risk of poor air quality?

SELECT
    Population_Density,
    Proximity_to_Industrial_Areas,
    COUNT(*) AS Poor_Quality_Count
FROM Pollution_Data
WHERE Air_Quality = 'Poor'
GROUP BY Population_Density, Proximity_to_Industrial_Areas
ORDER BY Poor_Quality_Count DESC;
```

population_density integer	proximity_to_industrial_areas double precision	poor_quality_count bigint
694	3.6	2
639	4.6	2
477	4.1	2
658	4.2	2
515	3.6	2
609	3.6	2
643	9	2
709	4.6	2
771	8.2	2
479	5.4	2
623	4.2	2
616	4.3	2
623	4	2
669	3.9	2
549	5.4	2
617	5.2	2
722	4.1	2
484	6.7	2
749	3.9	2
424	4.4	2

Q.16 WHAT PERCENTAGE OF RECORDS SHOW "GOOD" AIR QUALITY IN AREAS WITH PROXIMITY TO INDUSTRIAL ZONES?

```sql
--Q.16 What percentage of records show "Good" air quality in areas with proximity to industrial zones?

SELECT
    COUNT(*) * 100.0 / (SELECT COUNT(*) FROM Pollution_Data) AS Good_No_Industrial_Percentage
FROM Pollution_Data
WHERE Air_Quality = 'Good' AND Proximity_to_Industrial_Areas =
(SELECT MIN(Proximity_to_Industrial_Areas) FROM Pollution_Data);
```

good_no_industrial_percentage numeric
0.00000000000000000000

"AIR QUALITY NEAR INDUSTRIAL ZONES (≤ 10 KM)"

```sql
--"Air Quality Near Industrial Zones (≤10 km)"

SELECT
    COUNT(*) * 100.0 / (SELECT COUNT(*) FROM Pollution_Data) AS Good_Nearest_Industrial_Percentage
FROM Pollution_Data
WHERE Air_Quality = 'Good'
      AND Proximity_to_Industrial_Areas <= 10
```

good_nearest_industrial_percentage numeric
0.68054443554843875100

CONCLUSION

The analysis of the pollution dataset highlights critical insights into air quality and its influencing factors. Key findings include average pollutant levels, temperature, humidity distribution, and the categorization of air quality across regions. The data reveals significant correlations, such as the impact of industrial proximity on pollutant concentrations and the influence of population density on air quality.

CONCLUSION

Intermediate and advanced questions shed light on complex relationships, including the combined effect of multiple pollutants exceeding safe thresholds and the demographic factors contributing to poor air quality. These insights offer a deeper understanding of environmental dynamics and identify areas at higher risk for health concerns due to poor air quality.

CONCLUSION

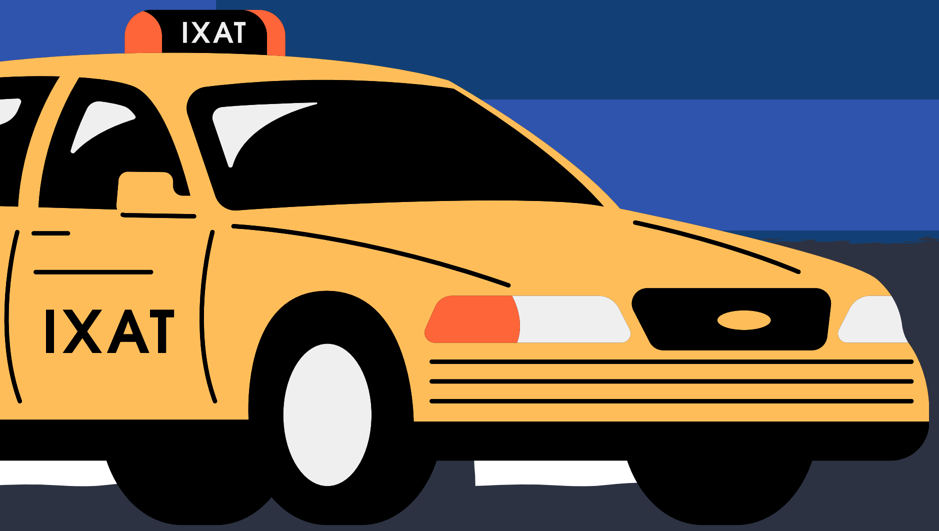
This project demonstrates the importance of data-driven decision-making in environmental management. By addressing these questions, stakeholders can prioritize pollution control measures, develop targeted policies, and improve public awareness, ultimately contributing to healthier living conditions.

PREVENTION METHODS FOR IMPROVING AIR QUALITY

Every individual has an important role in efforts to reduce air pollution. Here are some steps we can take as individuals:

Regulate Industrial Emissions

- Enforce strict emission standards for industries located near residential areas.
- Promote the use of cleaner technologies and regular maintenance to reduce sulfur dioxide (SO_2) and carbon monoxide (CO) emissions.

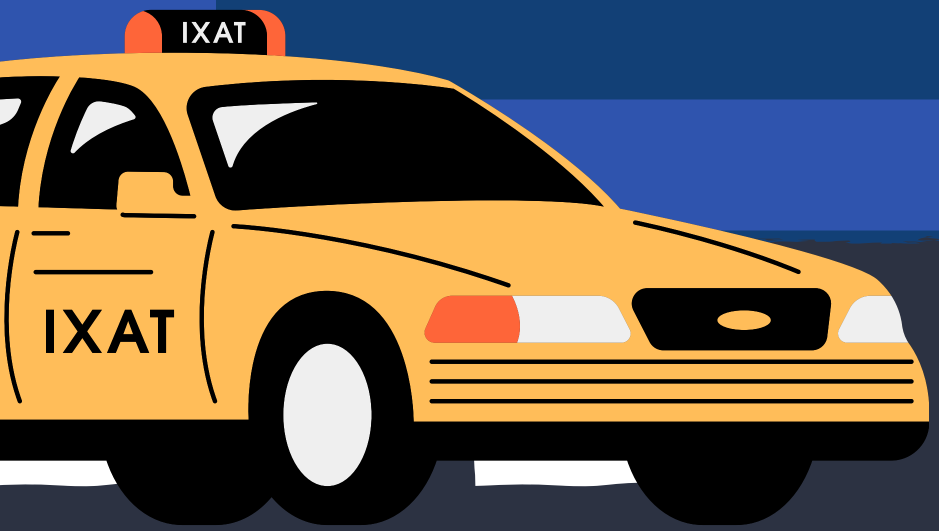


PREVENTION METHODS FOR IMPROVING AIR QUALITY

Every individual has an important role in efforts to reduce air pollution. Here are some steps we can take as individuals:

Promote Green Energy

- Transition to renewable energy sources like solar, wind, and hydropower to reduce reliance on fossil fuels.
- Encourage the adoption of electric vehicles (EVs) and provide subsidies for EV infrastructure.

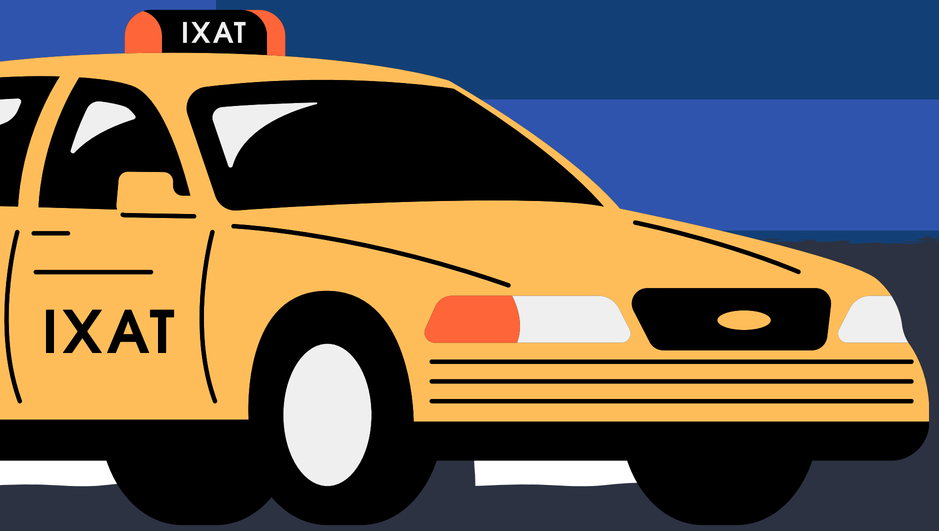


PREVENTION METHODS FOR IMPROVING AIR QUALITY

Every individual has an important role in efforts to reduce air pollution. Here are some steps we can take as individuals:

Enhance Public Transportation

- Develop and improve public transport systems to reduce individual car usage.
- Introduce low-emission zones and carpooling incentives to lower traffic-related pollution.

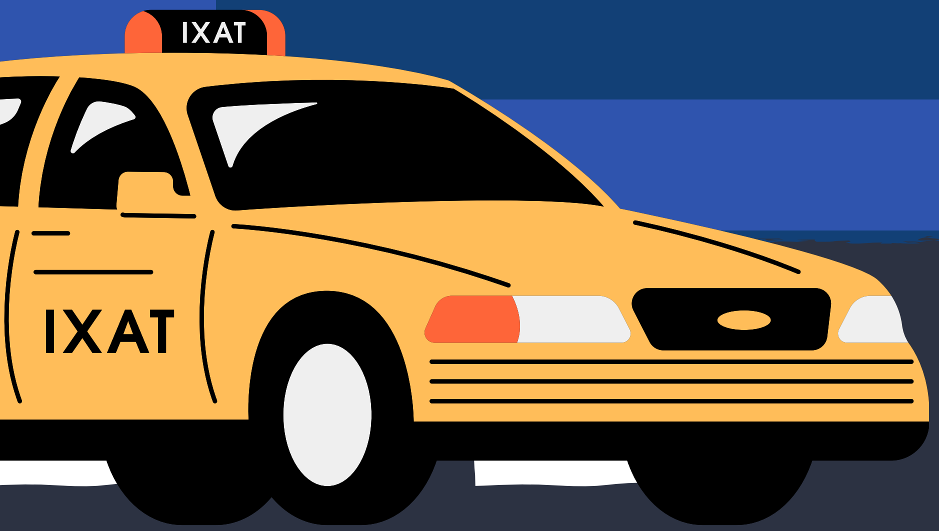


PREVENTION METHODS FOR IMPROVING AIR QUALITY

Every individual has an important role in efforts to reduce air pollution. Here are some steps we can take as individuals:

Reforestation and Urban Green Spaces

- Increase tree plantation drives to absorb particulate matter (PM2.5 and PM10) and other harmful gases.
- Establish urban parks to improve air quality and provide natural air filtration.

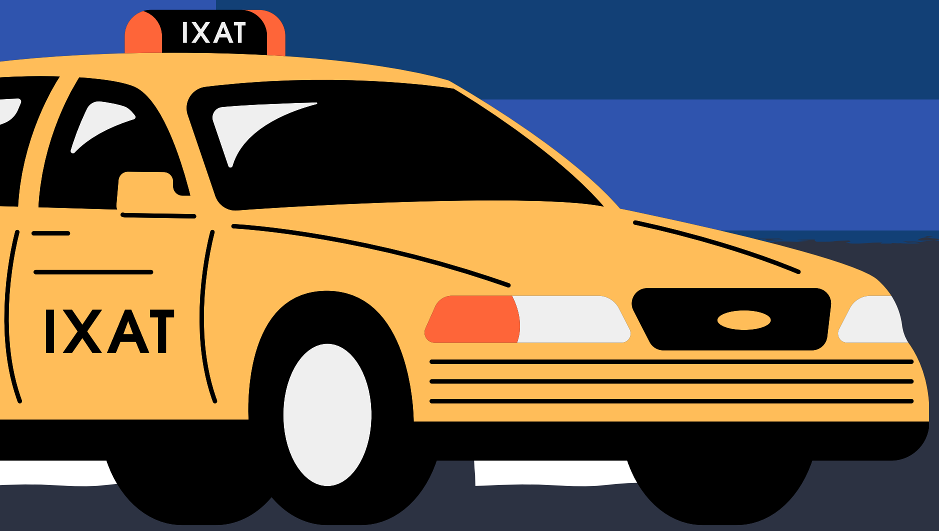


PREVENTION METHODS FOR IMPROVING AIR QUALITY

Every individual has an important role in efforts to reduce air pollution. Here are some steps we can take as individuals:

Monitor and Manage Construction Activities

- Implement strict regulations for dust control at construction sites.
- Use water sprinkling and dust barriers to reduce PM10 levels during construction.

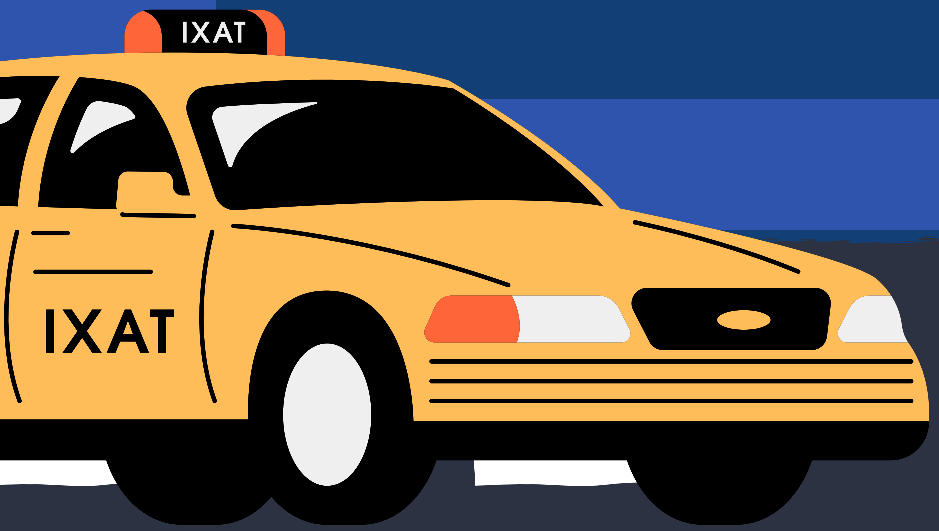


PREVENTION METHODS FOR IMPROVING AIR QUALITY

Every individual has an important role in efforts to reduce air pollution. Here are some steps we can take as individuals:

Public Awareness Campaigns

- Educate communities about the harmful effects of pollutants and ways to minimize exposure.
- Encourage individuals to adopt eco-friendly practices, such as reducing waste burning and using energy-efficient appliances.

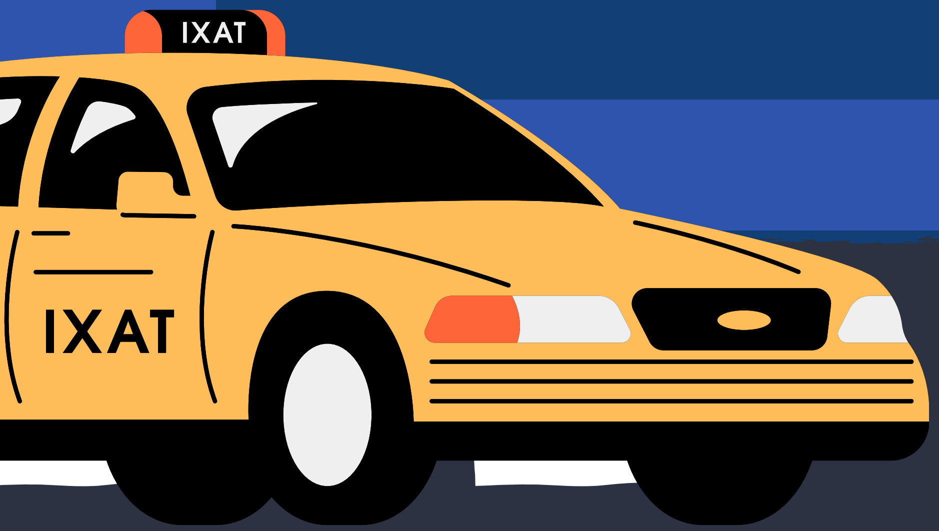


PREVENTION METHODS FOR IMPROVING AIR QUALITY

Every individual has an important role in efforts to reduce air pollution. Here are some steps we can take as individuals:

Strengthen Monitoring Systems

- **Set** up more air quality monitoring stations to identify high-risk areas.
- Use real-time data to implement timely interventions, such as temporary restrictions on industrial activity or traffic.



PREVENTION METHODS FOR IMPROVING AIR QUALITY

Every individual has an important role in efforts to reduce air pollution. Here are some steps we can take as individuals:

Encourage Research and Innovation

- **Supp**ort research into advanced pollution control technologies.
- Invest in air purifiers and filtration systems for both public and private spaces.

- **By implementing these prevention methods, governments, organizations, and individuals can collectively work toward reducing air pollution and creating healthier environments.**

SOURCES OF AIR POLLUTION

Air pollution comes from a variety of different sources, both natural and man-made. Here are some of the main sources of air pollution:

SICK OF POLLUTION

TRANSPORTATION

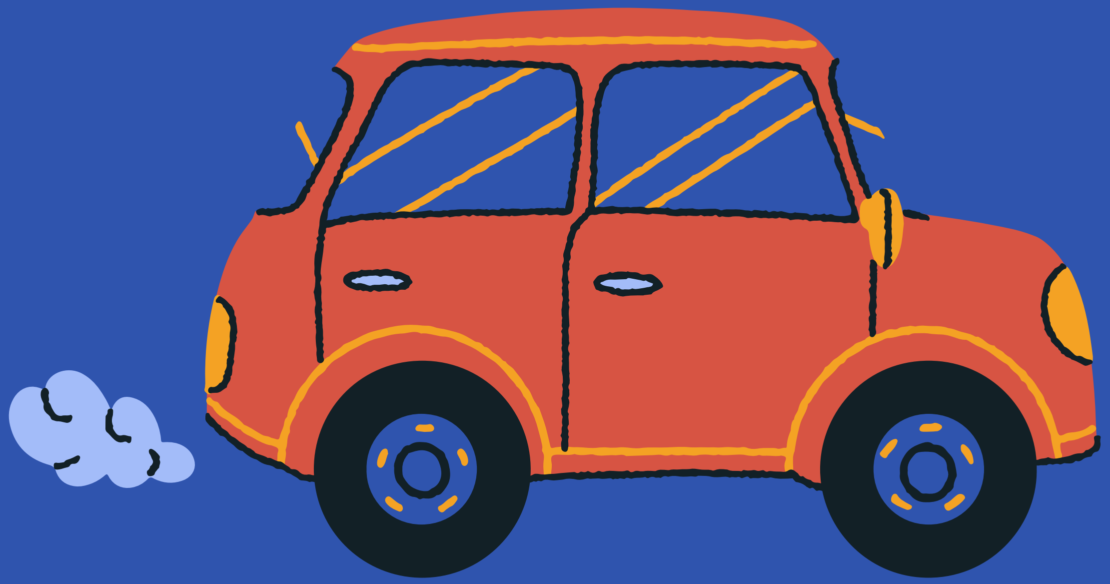
INDUSTRY

AGRICULTURE ACTIVITIES

INDOOR POLLUTION

AIR POLLUTION REDUCTION

There are several steps we can take to reduce air pollution and create a cleaner environment. Here are some solutions that can be implemented:



USES RENEWABLE ENERGY

REDUCING MOTOR VEHICLE USE

PLANTING TREES

SHARED RESPONSIBILITY

To overcome the problem of air pollution, joint responsibility is needed from all parties. This includes:

GOVERNMENT

Implement policies that support emission reduction and environmental protection.

SOCIETY

Adopt a more sustainable lifestyle, such as using public transport or cycling.

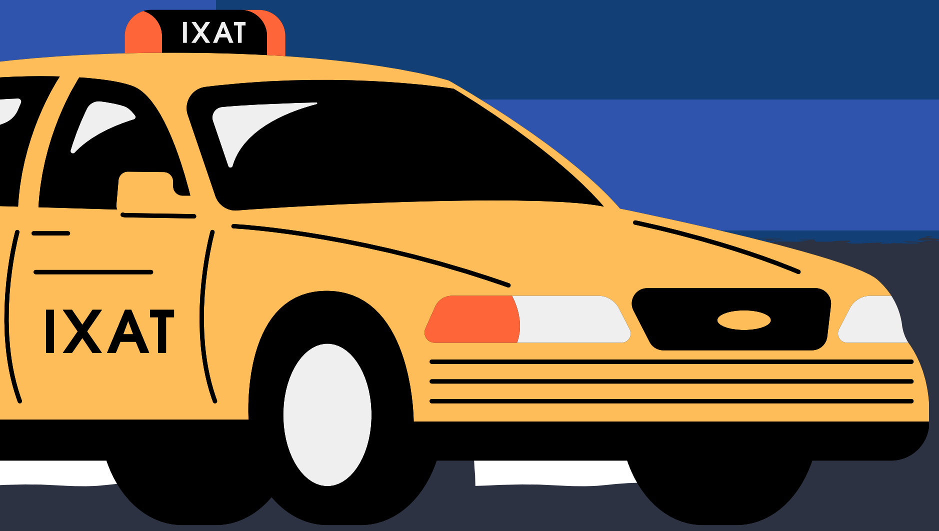
INDIVIDUAL

Reducing personal vehicle fuel use and choosing environmentally friendly alternatives.

INDIVIDUAL ROLES

Every individual has an important role in efforts to reduce air pollution. Here are some steps we can take as individuals:

- Reducing the Use of Private Vehicles
- Using Environmentally Friendly Products
- Increase Environmental Awareness
- Supports Environmentally Friendly Technology



CONCRETE ACTIONS

To overcome the problem of air pollution, joint responsibility is needed from all parties. This includes:

Some things you can do are as follows.

- Use Public Transportation or Bike
- Waste Recycling
- Use Eco Friendly Products
- Support Tree Planting

CONCLUSION

Maintaining pollution-free air is our collective responsibility to create a healthy environment for future generations. Through cooperation between governments, companies, communities and individuals, we can achieve significant change in reducing air pollution.

LET'S WORK TOGETHER TO MAINTAIN POLLUTION FREE AIR

FOR A BETTER FUTURE

THANK YOU

SICK OF
POLLUTION

LINKEDIN PROFILE