AIR QUALITY INDEX

A MINI-PROJECT REPORT

Submitted by

SABARIPRABU.M 240701446

SAKTHIBALASUNDARAM M 240701460

SABITHA.R 240701448

DHARSHINI.G 240701173

in partial fulfillment of the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

An Autonomous Institute

CHENNAI

NOVEMBER 2025

BONAFIDE CERTIFICATE

Certified that this project **"AIR QUALITY INDEX"** is theBonafide work of **"SABARIPRABU M,SAKTHIBALASUNDARAM M, SABITHA R,DHARSHINI G"** who carried out the project work under my supervision.

SIGNATURE

Mrs.M. BHAVANI

ASSISTANT PROFESSOR SG

Dept. of Computer Science and Engg,
Rajalakshmi Engineering College
Chennai

This mini project report is submitted for the viva voce examination to be held on

_____

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ABSTRACT

The Air Quality Index (AQI) project focuses on analyzing and understanding the quality of air in different regions using data science techniques. Air pollution has become a major environmental problem that affects human health and the planet. The main objective of this project is to study how different pollutants—such as PM2.5, PM10, CO, NO_2, and O_3—affect the air quality. The dataset is collected from reliable environmental monitoring sources, and data preprocessing techniques such as handling missing values, normalization, and visualization are applied to prepare the data for analysis. Various data science methods, including statistical analysis and trend visualization, are used to identify pollution patterns across time and location. The project also explores how weather conditions and traffic levels influence air pollution. The results help in understanding pollution hotspots and seasonal variations in air quality. Overall, this project demonstrates how data science can be applied to monitor environmental conditions, raise public awareness, and support government decisions for cleaner and healthier living environments.

ACKNOWLEDGEMENT

We express our sincere thanks to our beloved and honorable chairman **MR. S. MEGANATHAN** and the chairperson **DR. M.THANGAM MEGANATHAN** for their timely support and encouragement.

We are greatly indebted to our respected and honorable principal **Dr. S.N. MURUGESAN** for his able support and guidance.

No words of gratitude will suffice for the unquestioning support extended to us by our Head of The Department **Dr. E.M. MALATHY** and our **PROFESSOR Dr. AKSHAIKHAANTH** for being ever supporting force during our project work

We also extend our sincere and hearty thanks to our internal guide **Dr. AKSHAIKHAANTH,** for her valuable guidance and motivation during the completion of this project.

Our sincere thanks to our family members, friends and other staff members of computer science engineering.

1. ASWIN BALAJI P

2.HARISH S

3. LAL KRISHNA JM

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

The Air Quality Index (AQI) project focuses on analyzing air pollution levels using data science techniques. It helps identify major pollutants, study their effects on health, and understand patterns in air quality to promote cleaner and safer environments.

1.2 SCOPE OF THE WORK

The scope of this project includes collecting and analyzing air quality data from different regions to evaluate pollution levels. It involves identifying key pollutants, studying their variations over time, and determining factors affecting air quality. The project also uses data visualization and statistical techniques to present results clearly, helping authorities and the public take informed actions toward improving environmental health.

1.3 PROBLEM STATEMENT

Air pollution severely affects human health and the environment. This project aims to analyze air quality data to identify pollution levels and provide insights for improving environmental conditions.

1.4 AIM AND OBJECTIVES OF THE PROJECT

The aim of this project is to analyze air pollution data using data science methods. Its objectives include identifying key pollutants, studying pollution patterns, and providing insights to improve air quality and public awareness.

CHAPTER 2

SYSTEM SPECIFICATIONS

2.1 HARDWARE SPECIFICATIONS

The project requires a computer system with at least an Intel i3 or higher processor, 4GB RAM or more, 250GB hard disk storage, and a stable internet connection for data collection, analysis, and visualization tasks.

2.2 SOFTWARE SPECIFICATIONS

The project uses Windows or Linux operating system, Python programming language, Jupyter Notebook for implementation, and essential libraries such as NumPy, Pandas, Matplotlib, and Scikit-learn for data analysis and visualization.

CHAPTER 3

MODULE DESCRIPTION

The Air Quality Index (AQI) project is divided into several modules that work together to collect, process, analyze, and visualize air quality data effectively.

1.Data Collection and Preprocessing:

This module involves gathering air quality data from reliable sources and handling missing values. The data is cleaned, normalized, and formatted for analysis to ensure accuracy and consistency.

2. Exploratory Data Analysis (EDA):

In this module, visualizations are created using libraries like Matplotlib and Seaborn to study pollution trends, seasonal variations, and relationships between pollutants such as PM2.5, PM10, CO, and NO_2.

3. Model Building:

This module applies statistical and analytical techniques to identify patterns and correlations among pollutants and environmental factors like temperature and humidity. Predictive models may be used to forecast air quality levels.

4. Model Evaluation:

The analyzed data is presented using graphs, charts, and AQI category visualizations to make findings easy to interpret and informative for users.

5. Prediction Module:

This final module summarizes results, highlights pollution sources, and provides recommendations for improving air quality and raising public awareness.

CHAPTER 4

CODE

```
import pandas as pd


# Load dataset

data = pd.read_csv("city_day.csv")

data["Date"] = pd.to_datetime(data["Date"], errors='coerce')

data = data.dropna(subset=["AQI"])


# Compute daily average AQI

daily_avg = data.groupby(["Date",
"City"])["AQI"].mean().reset_index()


# Define alert system

def aqi_alert(aqi):

    if aqi > 300:

        return "Hazardous 😷"
```

```
    elif aqi > 200:
        return "Very Unhealthy 😷"
    elif aqi > 150:
        return "Unhealthy ⚠️"
    elif aqi > 100:
        return "Moderate ⚠️"
    else:
        return "Good 😊"

# Start continuous input loop
print("🌫️ Daily Air Quality Alert System")
print("Type 'exit' to quit.\n")

while True:
    city = input("Enter city name: ").strip().title()
    if city.lower() == "exit":
```

```
        print("\n👋 Exiting system. Stay safe!")
        break

    if city not in daily_avg["City"].unique():
        print(f" ❌ Sorry, {city} not found in dataset.\n")
        continue

    city_data = daily_avg[daily_avg["City"] ==
city].sort_values("Date")
    latest_date = city_data["Date"].iloc[-1]
    latest_aqi = city_data["AQI"].iloc[-1]
    alert = aqi_alert(latest_aqi)

    print("\n--- Air Quality Report ---")
    print(f"City: {city}")
    print(f"Date: {latest_date.date()}")
    print(f"AQI: {round(latest_aqi, 1)}")
```

```
print(f"Alert Level: {alert}")



# Add weather-style message
if alert.startswith("Good"):
    weather = "🌞 Clear skies, safe to go out."
elif alert.startswith("Moderate"):
    weather = "🌤️ Slight haze, sensitive groups take care."
elif alert.startswith("Unhealthy"):
    weather = "🌫️ Poor air quality — wear a mask outside."
else:
    weather = "☠️ Very poor — avoid outdoor activity!"



print("Weather Advice:", weather)
print("-" * 40)
```

CHAPTER 5

SCREEN SHOTS

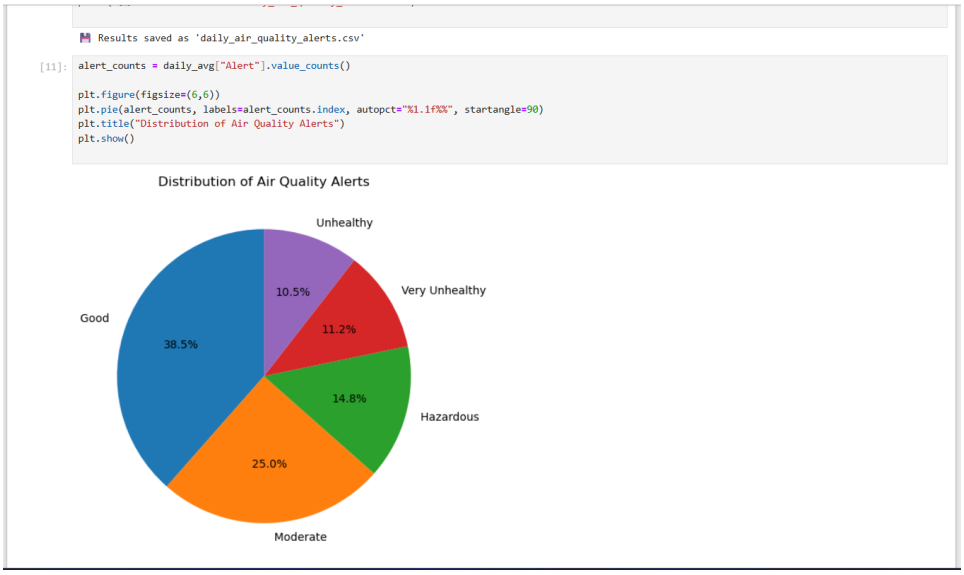
Delimiter: ,

	Date	City	Average_AQI	Alert
1	2015-01-01	Delhi	472.0	Hazardous
2	2015-01-02	Delhi	454.0	Hazardous
3	2015-01-03	Delhi	143.0	Moderate
4	2015-01-04	Delhi	319.0	Hazardous
5	2015-01-05	Delhi	325.0	Hazardous
6	2015-01-06	Delhi	318.0	Hazardous
7	2015-01-07	Delhi	353.0	Hazardous
8	2015-01-08	Delhi	383.0	Hazardous
9	2015-01-09	Delhi	375.0	Hazardous
10	2015-01-10	Delhi	376.0	Hazardous
11	2015-01-11	Delhi	379.0	Hazardous
12	2015-01-12	Delhi	375.0	Hazardous
13	2015-01-13	Delhi	366.0	Hazardous
14	2015-01-14	Delhi	353.0	Hazardous
15	2015-01-15	Delhi	340.0	Hazardous
16	2015-01-16	Delhi	356.0	Hazardous
17	2015-01-17	Delhi	360.0	Hazardous
18	2015-01-18	Delhi	370.0	Hazardous
19	2015-01-19	Delhi	362.0	Hazardous
20	2015-01-20	Delhi	340.0	Hazardous
21	2015-01-21	Delhi	338.0	Hazardous
22	2015-01-22	Delhi	332.0	Hazardous
23	2015-01-23	Delhi	254.0	Very Unhealthy
24	2015-01-24	Delhi	324.0	Hazardous
25	2015-01-25	Delhi	333.0	Hazardous
26	2015-01-26	Delhi	292.0	Very Unhealthy
27	2015-01-27	Delhi	318.0	Hazardous

5.1 - Sample Dataset

```
[ ]: import pandas as pd

# Load dataset
data = pd.read_csv("city_day.csv")
data["Date"] = pd.to_datetime(data["Date"], errors='coerce')
data = data.dropna(subset=["AQI"])

# Compute daily average AQI
daily_avg = data.groupby(["Date", "City"])["AQI"].mean().reset_index()

# Define alert system
def aqi_alert(aqi):
    if aqi > 300:
        return "Hazardous 🚨"
    elif aqi > 200:
        return "Very Unhealthy 🚨"
    elif aqi > 150:
        return "Unhealthy ⚠️"
    elif aqi > 100:
        return "Moderate ⚠️"
    else:
        return "Good 😊"

# Start continuous input loop
print("🌫️ Daily Air Quality Alert System")
print("Type 'exit' to quit.\n")

while True:
    city = input("Enter city name: ").strip().title()
    if city.lower() == "exit":
        print("\n👋 Exiting system. Stay safe!")
        break

    if city not in daily_avg["City"].unique():
        print(f"❌ Sorry, {city} not found in dataset.\n")
        continue

    city_data = daily_avg[daily_avg["City"] == city].sort_values("Date")
    latest_data = city_data["Date"].iloc[-1]
```

5.2 - Model Prediction AQI



5.3 – PIE CHART

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

The Air Quality Index (AQI) project successfully demonstrates how data science techniques can be applied to analyze and understand air pollution levels. By collecting, cleaning, and visualizing air quality data, the project helps identify major pollutants, their trends, and their impact on human health and the environment. The insights gained can assist in promoting awareness and supporting government policies for cleaner air.

In the future, this project can be enhanced by integrating real-time air quality monitoring systems, using advanced machine learning models for prediction, and developing a web or mobile application to display live AQI updates. These improvements will make the system more interactive, accurate, and useful for the public and environmental agencies.

REFERENCES

1. https://www.kaggle.com/blastchar/telco-customer-churn
2. https://scikit-learn.org/
3. https://pandas.pydata.org/
4. https://numpy.org/
5. https://matplotlib.org/
6. https://seaborn.pydata.org/