A Project Report

On

Air Quality Management and Predicting Data Analysis

Submitted by:

Kunal Kaushik

Under the guidance of

Mr. Prabhat Das

(PGT-IP)

ARMY PUBLIC SCHOOL JORHAT

CHARAIBAHI, MILITARY STATION

Assam-785616

CERTIFICATE BY PRINCIPAL

This is to certify that this project report entitled “Air Quality Management and Predicting Data Analysis” submitted by Kunal Kaushik to Army Public School Jorhat has been examined and evaluated.

The report has been prepared as per the regulations of CBSE and qualifies to be accepted.

Date:

Place:

Mrs Firdausi Sultana Hazarika
(Principal)
Army Public School Jorhat

CERTIFICATE BY EXAMINERS

This is to certify that this project report entitled "Air Quality Management and Predicting Data Analysis" is the bona fide work of Kunal Kaushik who carried out the project work under my supervision and guidance.

To the best of my knowledge, the matter embodied in the report has not been submitted to any other institute for the award of any other degree.

Date:

Place:

Prabhat das

(External Examiner) (Internal Examiner)

ACKNOWLEDGEMENT

I take this opportunity to extend my heart full gratitude to Army Public School Jorhat for providing me the opportunity.

I am highly grateful to my guide Mr. Prabhat Das, PGT-IP, Army Public School Jorhat for giving us the opportunity to work under him and providing us an ample guidance and support through the project.

Lastly, I would also like to thank the authors whose publications guided us regarding our project.

DECLARATION

I admit that this report is of my own work and all the sources of the information used in this report have fully acknowledged.

I hereby declare that the dissertation work entitled "AIR QUALITY MANAGEMENT AND PREDICTING DATA ANALYSIS" submitted to the Army Public School Jorhat, is prepared by me and was not submitted to any other institution for award of any other degree.

Date:

Place:

Signature

Abstract

In the era of today's world, where the globe is increasing it's fecundity monumentally, the introduction of computers has contributed to a boundless scale to this everlasting race. The invention of computers and mechanisation has eased the human difficulties, where one had to store all the data in the form of physical means, adding to the efforts, space and expenditure.

In this project, we're going to introduce you to a data science project on Air Quality Monitoring and Predicting data analysis. The user here inputs their e-mail, which then receives an OTP through the program on the mail itself' asking the user to authenticate the code. On the successful authentication if the code , the user is provided with the list of options, showing them a list of options in the program. The user is to select the option(s) they wish to see from the given list by feeding in the numbers allotted to them. On feeding in the value(s), they are displayed with the information stored in the dataset in the form of Figures of graphs.

The dataset we used for the Air Quality Management and Predicting data analysis task consists of PM2.5, NO, NO2, CO, SO2, O3, Benzene, Toluene, Xylene and AQI. The dataset is provided by Kaggle which allows users to find and publish data sets, explore and build models in a web-based data science environment.

CONTENTS

	Page No.
1. Tools and libraries used	1
2. Introduction	2
3. Project Overview	3
4. Source code	11
5. Commands used in MySQL	16
6. Conclusion and future work	17
7. References	18

List of Figures

Page No.

- Fig 1. MySql Table 3
- Fig 2.1. Login for registered user. 3
- Fig 2.2. Login for new user. 3
- Fig 3. Screen after login. 4
- Fig 4. Output for option 1. 4
- Fig 5. Output for option 2. 4
- Fig 6. Output for option 3. 5
- Fig 7.1. Output for option 4.1. 5
- Fig 7.2. Output for option 4.2. 6
- Fig 7.3. Output for option 4.3. 6
- Fig 7.4. Output for option 4.4. 7
- Fig 7.5. Output for option 4.5. 7
- Fig 7.6. Output for option 4.6. 8
- Fig 7.7. Output for option 4.7. 8
- Fig 7.8. Output for option 4.8. 9
- Fig 7.9. Output for option 4.9. 9
- Fig 7.10. Output for option 4.10. 10

Tools and Libraries Used

MySQL

MySQL is an open-source relational database management system (RDBMS). Its name is a combination of "My", the name of co-founder Michael Widenius's daughter, and "SQL", the abbreviation for Structured Query Language. A relational database organizes data into one or more data tables in which data types may be related to each other; these relations help structure the data. SQL is a language, programmers use to create, modify and extract data from the relational database, as well as control user access to the database. In addition to relational databases and SQL, an RDBMS like MySQL works with an operating system to implement a relational database in a computer's storage system, manages users, allows for network access and facilitates testing database integrity and creation of backups.

MySQL is free and open-source software under the terms of the GNU General Public License, and is also available under a variety of proprietary licenses. MySQL was owned and sponsored by the Swedish company MySQL AB, which was bought by Sun Microsystems (now Oracle Corporation). In 2010, when Oracle acquired Sun, Widenius forked the open-source MySQL project to create MariaDB [1].

PyCharm

PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as data science with Anaconda.

PyCharm is cross-platform, with Windows, macOS and Linux versions. The Community Edition is released under the Apache License, and there is also Professional Edition with extra features – released under a proprietary license [2].

Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged. SciPy makes use of Matplotlib.

Matplotlib was originally written by John D. Hunter. Since then it has an active development community and is distributed under a BSD-style license [3].

Introduction

The starting point of air quality monitoring is to first study if an area has an air pollution problem. Monitoring helps in assessing the level of pollution in relation to the ambient air quality standards. Standards are a regulatory measure to set the target for pollution reduction and achieve clean air. Robust monitoring helps to guard against extreme events by alerting people and initiate action. We regulate a total of 10 pollutants, including PM2.5 (particulate matter of up to 10 micron and up to 2.5micron size), NO, NO2, CO, SO2, O3, Benzene, Toluene, Xylene and AQI.

Across cities, only SO2 and NO2 are monitored regularly. Other pollutants, such as PM2.5, O3, CO are monitored in select cities as capacity is still being built. India has set a target for states to meet National Ambient Air Quality Standards (NAAQS) in urban areas by 2017.

The following program is an attempt to develop a program that can analyse the data of Air quality from a given dataset. It includes a simple registration system where the user can register him/her into the program with the help of a one-time password (OTP) sent to the email address used for the registration process. The program then gives us a number of options to get different information about the category entered present in the given dataset. A graph is also generated using the Matplotlib library that gives us a visual representation of the polarity of the given data.

Project Overview

```
mysql> desc user_id;
+--------+-------------+------+-----+---------+--------+
| Field  | Type        | Null | Key | Default | Extra  |
+--------+-------------+------+-----+---------+--------+
| email  | varchar(30) | YES  |     | NULL    |        |
| name   | varchar(30) | YES  |     | NULL    |        |
| id     | int         | YES  |     | NULL    |        |
+--------+-------------+------+-----+---------+--------+
3 rows in set (0.00 sec)

mysql> select * from user_id;
+---------------------------+---------+------+
| email                     | name    | id   |
+---------------------------+---------+------+
| probalborah44@gmail.com | Probal |    1 |
+---------------------------+---------+------+
1 row in set (0.00 sec)
```

Figure 1: All the information regarding the users are stored in this table.

```
C:\Users\proba\PycharmProjects\pythonProject2\venv\Scripts\python.exe "C:/Users/proba/PycharmProjects/pythonProject2/venv/share/air quality monitoring and prediction syste
**********
Welcome to Air quality monitoring and ptrediction sysytem
**********
SELECT THE OPTION TO BEGIN !
Press 1 to Login
Press 2 to Register
1
ENTER THE USER ID TO LOGIN:
probalborah44@gmail.com
Please wait we are sending the OTP to the given ID........
Enter the OTP:
Login Successful !!!
```

Figure 2.1: When the user chooses option 1 the program asks for his/her email address, on entering the email id, if the email id exists then the login procedure is completed.

```
**********
Welcome to Air quality monitoring and ptrediction sysytem
**********
SELECT THE OPTION TO BEGIN !
Press 1 to Login
Press 2 to Register
2
ENTER YOUR NAME
probal
ENTER EMAIL ID
probalborah44gmail.com
User has been successfully registered!!!!
ENTER THE USER ID TO LOGIN:
probalborah44@gmail.com
Please wait we are sending the OTP to the given ID........
Enter the OTP:
278949
Login Successful !!!
```

Figure 2.2: Login procedure for new users.

Figure3: After the login procedure is completed, the user is presented with the following list of options.

Figure4: Output for option 1.

Figure5:Output for option 2.

```
Please follow the following instructions
**********************************************
Press 1 to see the highest value of any category
Press 2 to see the lowest value of any category
Press 3 to see the mean value of any category
Press 4 to generate the graph of the average of any category of different cities
**********************************************
Enter your response:3
Enter the Category:AQI
197.310002257846
```

Figure6: Output for option 3.

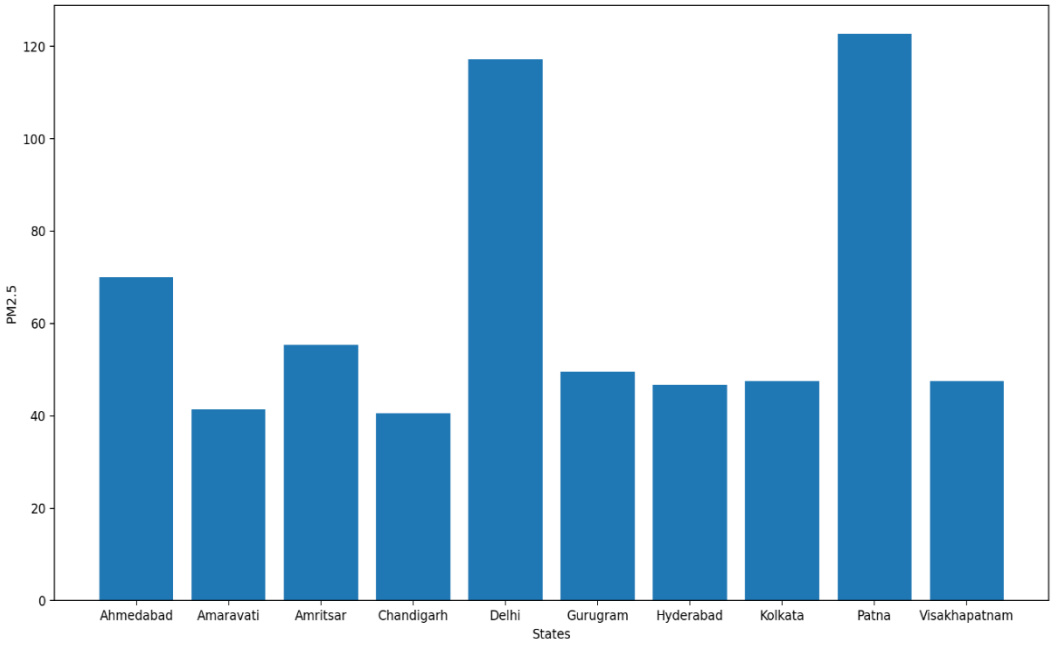

Figure7.1: Output for 4 when user enters PM2.5.

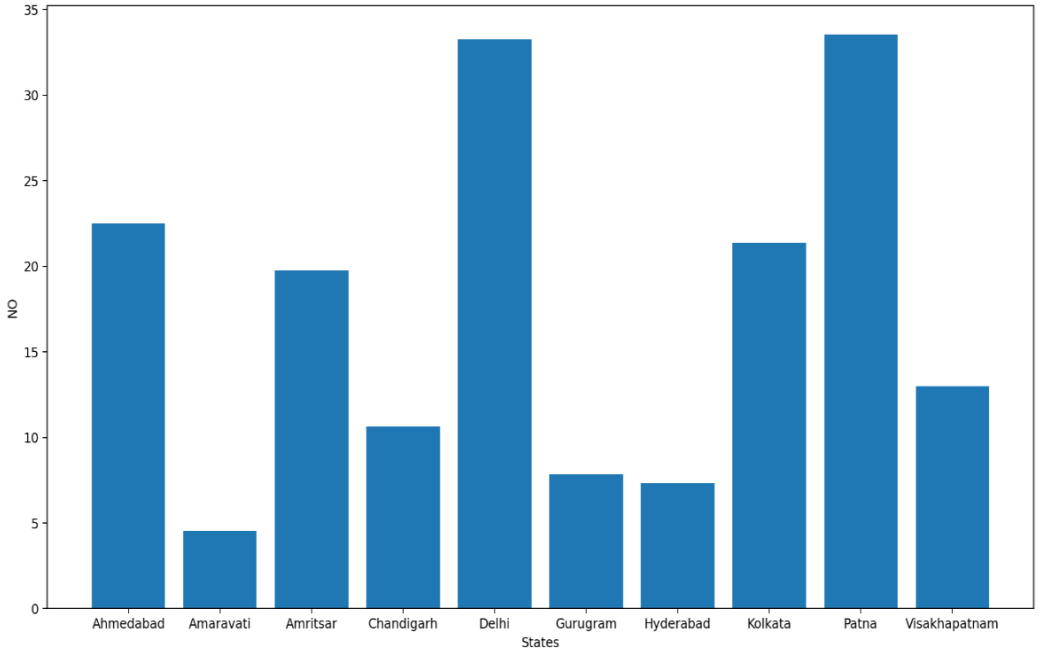

Figure7.2: Output for 4 when user enters NO.

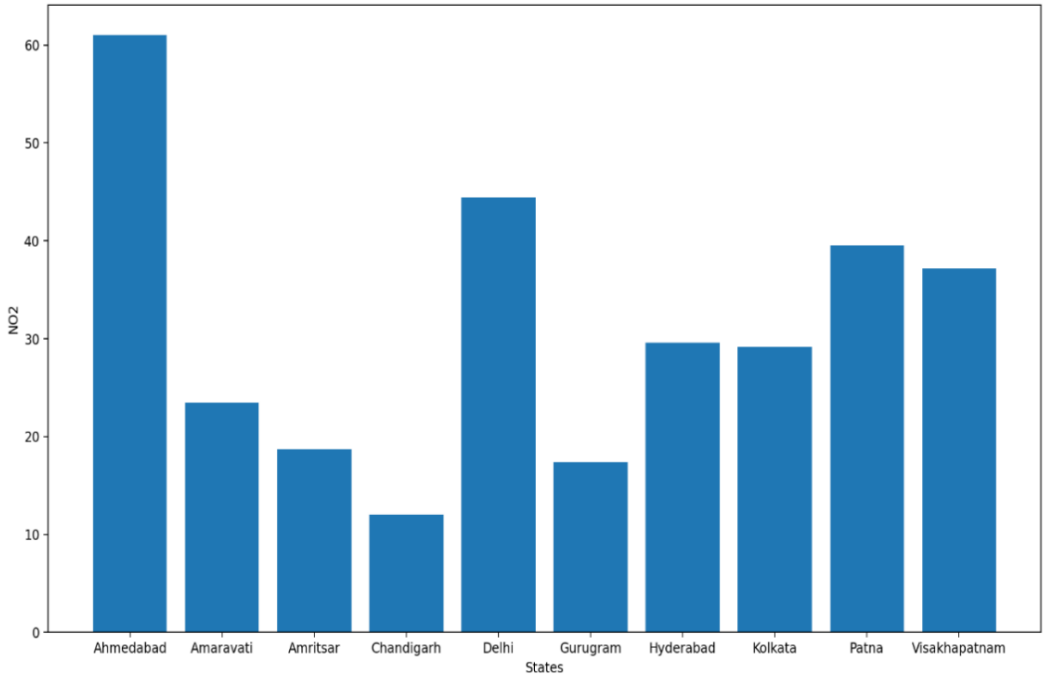

Figure7.3: Output for 4 when user enters NO2.

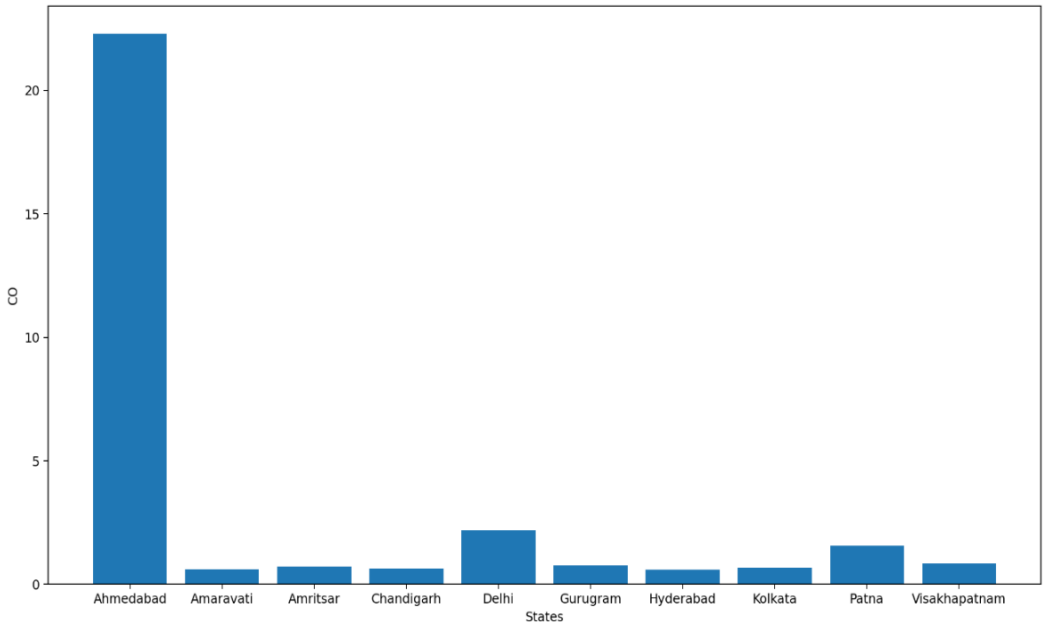

Figure7.4: Output for 4 when user enters CO.

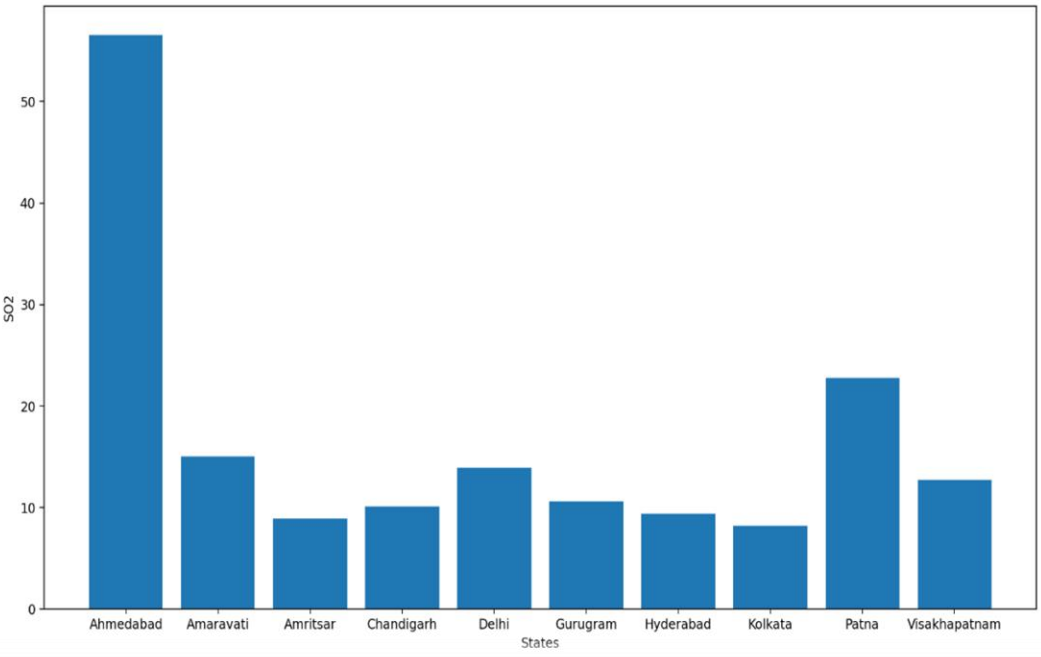

Figure7.5: Output for 4 when user enters SO2.

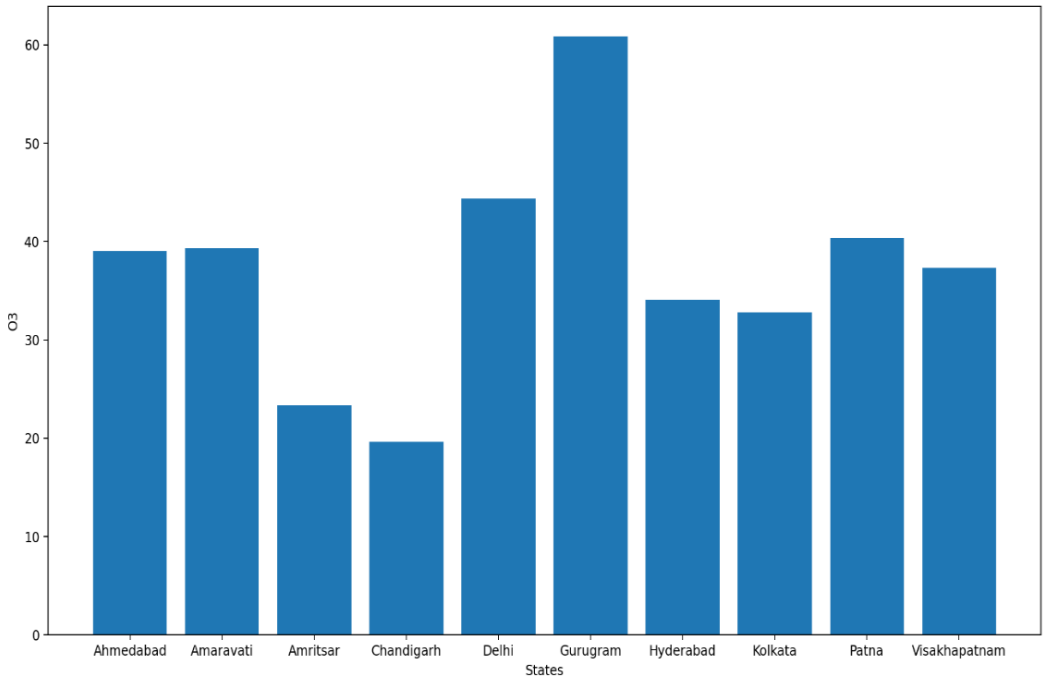

Figure7.6: Output for 4 when user enters O3.

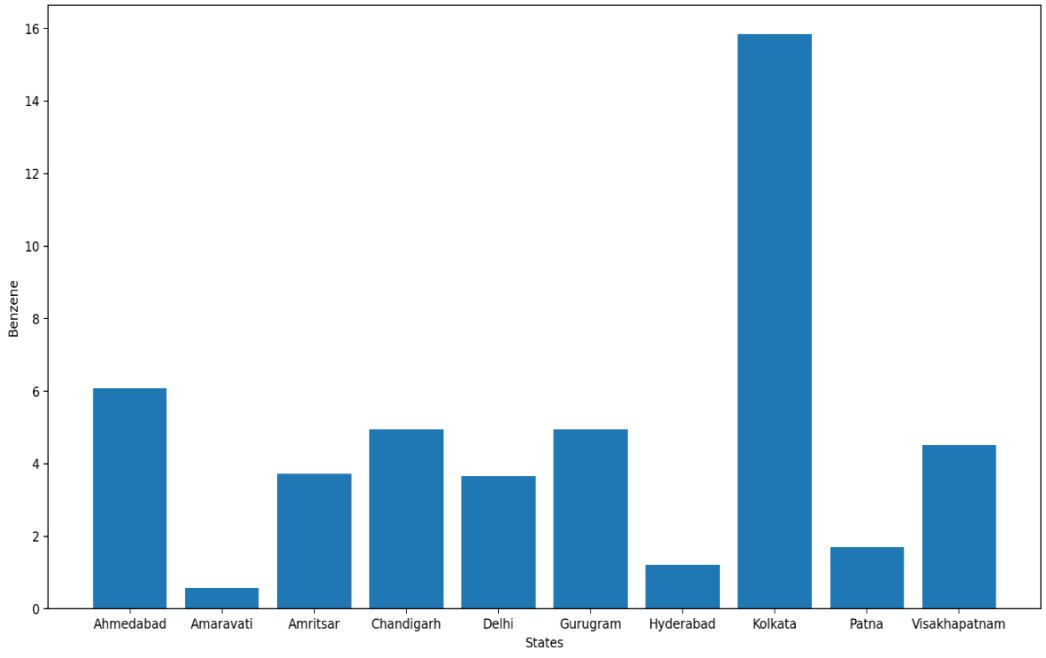

Figure7.7: Output for 4 when user enters Benzene.

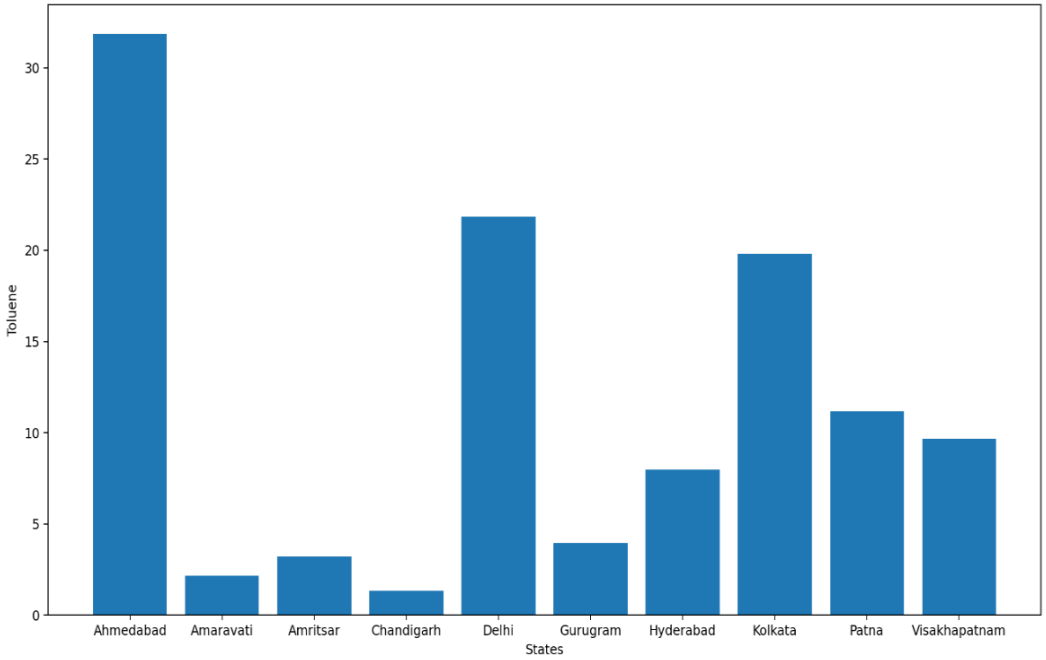

Figure7.8: Output for 4 when user enters Toluene.

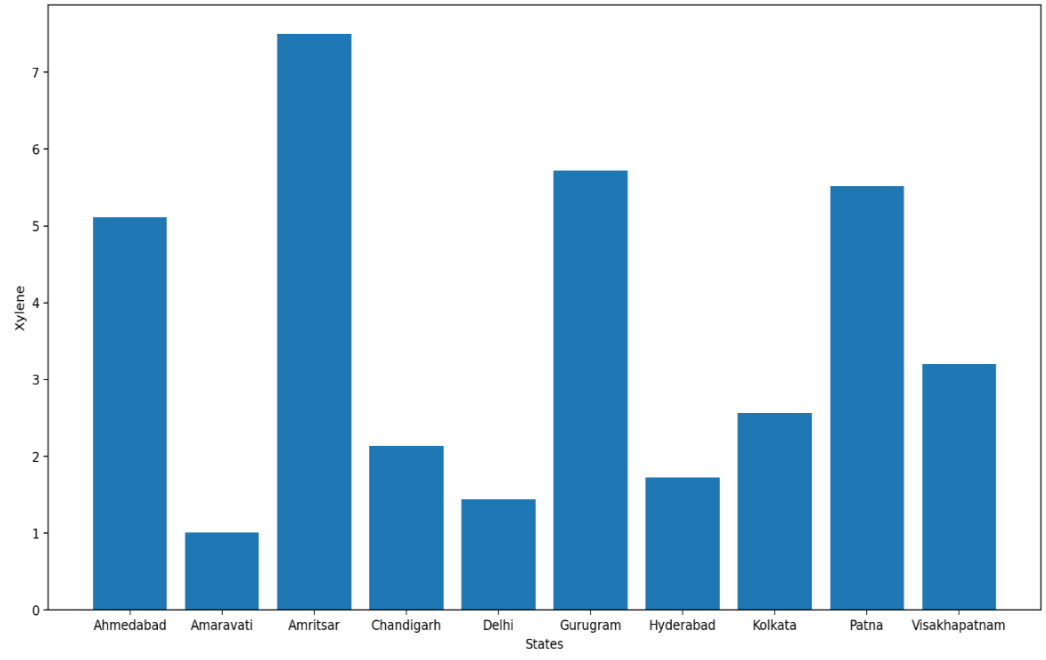

Figure7.9: Output for 4 when user enters Xylene.

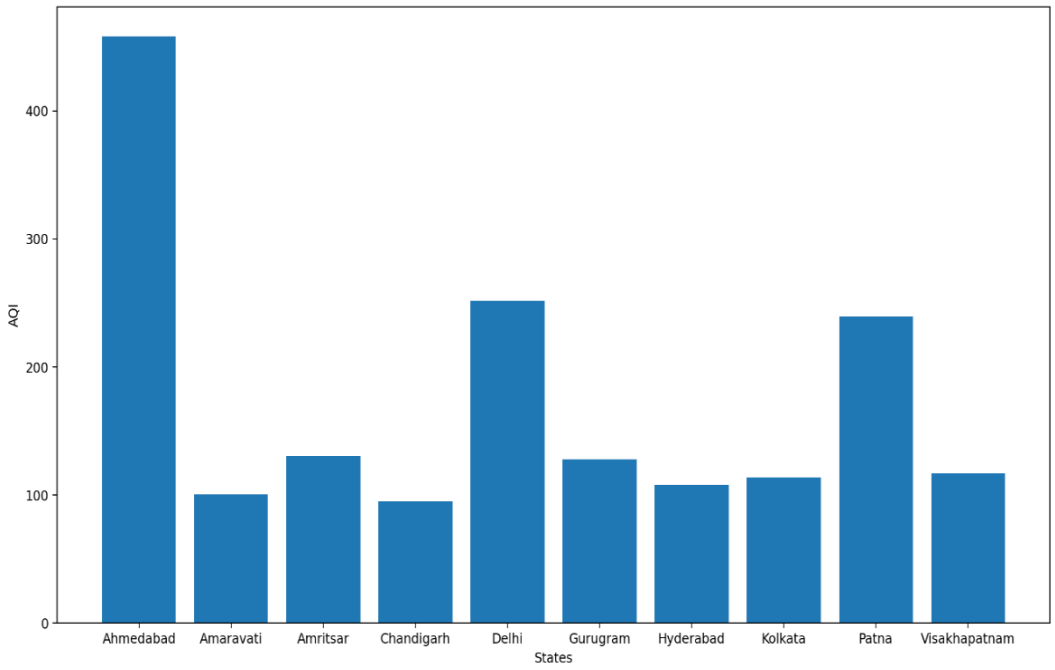

Figure7.10: Output for 4 when user enters AQI.

Source Code

MAIN

```
import connector as con
import matplotlib.pyplot as plt
import pandas as pd
import auth
def registration():
    print("ENTER YOUR NAME")
    name=input()
    print("ENTER EMAIL ID")
    email=input()
    query="insert into registration (name,email) values" +"('"+name+"','"+email+"');"
    con.cursor.execute(query)
    con.dbc.commit()
    print("User has been successfully registered!!!!")
    login()

def login():
    print("ENTER THE USER ID TO LOGIN: ")
    user_id = input()
    fetch_query = "select * from registration;"
    con.cursor.execute(fetch_query)
    count = 0
    for i in con.cursor:
        if user_id == i[2]:
            count = count + 1
            print("Please wait we are sending the OTP to the given ID.....")
            auth.auth(user_id)

    if count == 0:
```

```
        print("User not registered")
        registration()
    else:
        print("User Registered")
    import pro
    print("Do you want to go back?")
    print("type Yes to go back")
    print("type No to end")
    user_input = input()
    if user_input=="Yes":
        back()
    elif user_input=="No":
        exit()

def dashboard():
    edf = pd.read_csv("aqi.csv")
    print("*" * 46)
    print("Please follow the following instructions")
    print("*" * 46)
    print("Press 1 to see the highest value of any category")
    print("Press 2 to see the lowest value of any category")
    print("Press 3 to see the mean value of any category")
    print("Press 4 to generate the graph of the average of any category of different cities")
    print("*" * 46)
    r1 = int(input("Enter your response:"))
    # print(edf.to_markdown())
    if r1 == 1:
        cat = input("Enter the Category:")
        print(edf[cat].max())
    if r1 == 2:
```

```
        cat1 = input("Enter the Category:")
        print(edf[cat1].min())
    if r1 == 3:
        cat2 = input("Enter the Category:")
        print(edf[cat2].mean())
    if r1 == 4:
        a1 = input(
            "Enter the category of which you want to see the average(PM2.5,NO,NO2,CO,SO2,O3,Benzene,Toluene,Xylene,AQI):")
        a = edf[edf['City'] == 'Ahmedabad'][a1].mean()
        b = edf[edf['City'] == 'Amaravati'][a1].mean()
        c = edf[edf['City'] == 'Amritsar'][a1].mean()
        d = edf[edf['City'] == 'Chandigarh'][a1].mean()
        e = edf[edf['City'] == 'Delhi'][a1].mean()
        f = edf[edf['City'] == 'Gurugram'][a1].mean()
        g = edf[edf['City'] == 'Hyderabad'][a1].mean()
        h = edf[edf['City'] == 'Kolkata'][a1].mean()
        i = edf[edf['City'] == 'Patna'][a1].mean()
        j = edf[edf['City'] == 'Visakhapatnam'][a1].mean()
        data = [a, b, c, d, e, f, g, h, i, j]
        h = ["Ahmedabad", "Amaravati", "Amritsar", "Chandigarh", "Delhi", "Gurugram",
"Hyderabad", "Kolkata", "Patna",
             "Visakhapatnam"]
        plt.bar(h, data)
        plt.xlabel("States")
        plt.ylabel(a1)
        plt.show()
        dashboard()
print("*"*46)
print("Welcome to Air quality monitoring and prediction system")
print("*"*46)
```

```
print("SELECT THE OPTION TO BEGIN !")
print("Press 1 to Login")
print("Press 2 to Register")

user_input=int(input())
if user_input==1:
    login()
elif user_input==2:
registration()
```

AUTHENTICATION AND OTP SENDER

```
import smtplib
import random
def auth(user_id):
    s=smtplib.SMTP('smtp.gmail.com',587)
    s.starttls()
    s.login('group3@apsjorhat.org','apsj#12345678')
    otp=random.randint(111111,999999)
    message=str(otp)
    s.sendmail('group3@apsjorhat.org',user_id,message)
    s.quit()
    print('Enter the OTP:')
    val=int(input())
    if val==otp:
        print("Login Successful !!!")
    else:
        print("Incorrect OTP")
```

CONNECTOR

```
import mysql.connector as mc
try:
    dbc=mc.connect(host="localhost",user="root",passwd="hellyeah11",database="project")
    cursor=dbc.cursor()
except Exception as e:
    print(e)
```

DASHBOARD

```
import pandas as pd
import matplotlib.pyplot as plt
edf=pd.read_csv("aqi.csv")
print("Welcome to our system!!!")
```

```
print("Here you can monitor the quality of air bases on ten categories, that is
PM2.5,NO,NO2,CO,SO2,O3,Benzene,Toluene,Xylene and AQI.")
print("The data can be monitored of ten different cities in India")
print("*"*46)
print("Please follow the following instructions")
print("*"*46)
print("Press 1 to see the highest value of any category")
print("Press 2 to see the lowest value of any category")
print("Press 3 to see the mean value of any category")
print("Press 4 to generate the graph of the average of any category of different cities")
print("*"*46)
r1=int(input("Enter your response:"))
# print(edf.to_markdown())
if r1==1:
    cat=input("Enter the Category:")
    print(edf[cat].max())
elif r1==2:
    cat1=input("Enter the Category:")
    print(edf[cat1].min())
elif r1==3:
    cat2=input("Enter the Category:")
    print(edf[cat2].mean())
elif r1==4:
    a1=input("Enter the category of which you want to see the
average(PM2.5,NO,NO2,CO,SO2,O3,Benzene,Toluene,Xylene,AQI):")
    a=edf[edf['City']=='Ahmedabad'][a1].mean()
    b=edf[edf['City']=='Amaravati'][a1].mean()
    c=edf[edf['City']=='Amritsar'][a1].mean()
    d=edf[edf['City']=='Chandigarh'][a1].mean()
    e=edf[edf['City']=='Delhi'][a1].mean()
    f=edf[edf['City']=='Gurugram'][a1].mean()
    g=edf[edf['City']=='Hyderabad'][a1].mean()
    h=edf[edf['City']=='Kolkata'][a1].mean()
    i=edf[edf['City']=='Patna'][a1].mean()
    j=edf[edf['City']=='Visakhapatnam'][a1].mean()
    data=[a,b,c,d,e,f,g,h,i,j]

h=["Ahmedabad","Amaravati","Amritsar","Chandigarh","Delhi","Gurugram","Hyderabad","
Kolkata","Patna","Visakhapatnam"]
    plt.bar(h,data)
    plt.xlabel("States")
    plt.ylabel(a1)
    plt.show(
```

Commands Used In MySql

Creating Database

Create database project;

Using Database

Use project;

Creating Table and Inserting Values

Create table user_id(email varchar(30), name varchar(30), id int);

To Fetch All Values

Select * from user_id;

Conclusion and Future Work

In this project we have created a program that can be used to analyse the quality of air of 10 cities in India. Most of the world's data is unstructured and unorganised and a program such as ours can help sort the data in an efficient manner and generate useful information. Air Quality Monitoring And Predicting Data Analysis is important because it helps analyse the growth of pollutants in air.

The functions performed by this project program are in accordance to our assumptions for further upgradation, we can also add features such as a Graphical User Interface to this project, which can be developed in the form of a web, desktop or mobile application. This project can also be scaled to analyse global data in real time.

References

[1] https://en.wikipedia.org/wiki/MySQL

[2] https://en.wikipedia.org/wiki/PyCharm

[3] https://en.wikipedia.org/wiki/MatplotlibS