STUDY OF AIR QUALITY IN STATE OF BIHAR

Project Report Submitted To Central University Of South Bihar
In Partial Fulfillment Of The Requirement Of The Degree Of

MASTERS OF SCIENCE IN STATISTICS

Submitted by

SUSHANT SHEKHAR DASHPUTE

Enrollment No : CUSB2102212007

Session: 2021– 2023

Supervised by

Dr. SANDEEP KUMAR MAURYA

Assistant Professor

Department of Statistics
School of Mathematics, Statistics and Computer Science
Central University of South Bihar
Gaya, Bihar, India - 824236

DECLARATION

The work embodied in the project entitled as “STUDY OF AIR QUALITY IN STATE OF BIHAR” was initiated at the Department of Statistics under School of Mathematics, Statistics, and Computer science, Central University of South Bihar Gaya, Bihar, in partial fulfillment of the requirement for the award of M.Sc. Degree in Statistics. This work has not been submitted in part or full, to this or any other university or institution, for any degree or diploma.

Date………………. Signature

(SUSHANT SHEKHAR DASHPUTE)

CERTIFICATE

This is to certify that the project entitled "STUDY OF AIR QUALITY IN STATE OF BIHAR" is submitted by SUSHANT SHEKHAR DASHPUTE, enrollment number CUSB2102212007, in partial fulfillment for award of M.Sc. Degree in Statistics of Central university of South Bihar. This work has not been submitted for the award of any degree/diploma of this or any other university and it is his original work.

Date……………………

Signature

(Dr. Sandeep Kumar Maurya)

Supervisor

Date…………………….

Signature

(Dr. Sunit Kumar)

Head Of Department

ACKNOWLEDGEMENT

Every Project is a sum total of various factors like hard work, difficulties, dedications. For the Successful completion of a project always we will have some helping hands. Those helping hands can be either our teachers or our family. This project work would not have been possible in this structure without the priceless effort of many fellow beings. And I would like to take this opportunity to thank all the helping hands during my project work.

First and foremost, I would like to express my Special thanks of gratitude to my supervisor Dr. **Sandeep Kumar Maurya**, Department of Statistics, Central University of South Bihar, for his able guidance, valuable feedback and constant encouragement throughout the duration of the research. Working with him has been an extremely knowledgeable experience for me. My sincere thanks to Dr. Sunit Kumar, Head of Department, Department of Statistics who gave directions from start to end of the study. I would also like to extend my gratuity to other faculty members of Department of Statistics, Dr. Ashok Kumar, Associate professor, Dr. Richa Vatsa, Dr. Kamlesh Kumar, Assistant Professor for their support and guidance during these two years, who put their faith and urged me to do better. Last but not the least, I would like to extend my love and gratitude to each and every one who helped me to successfully complete this project work.

SUSHANT SHEKHAR DASHPUTE

Table of Contents

List Of Figures

List Of Tables

ABSTRACT

This research study aims to analyze the air quality of Bihar state, with a focus on identifying the most polluted cities, forecasting PM2.5 concentration in a polluted region using statistical tools such as trend method, smoothing method, and model fitting. Additionally, the study aims to examine the awareness of air pollution among people in Gaya city, determine its effect on human health, and identify which age group and working field are mostly affected by air pollution. The research utilized PM2.5 data from five air quality monitoring stations, with major findings highlighting the need for targeted interventions to reduce exposure to air pollution among individuals living in urban areas, working outdoors, and those who smoke. The study also found a statistically significant association between occupation and the extent to which respondents report being affected by air pollution. Overall, the study reveals the pressing need to address air pollution in Bihar to improve public health and the environment.

CHAPTER 1 : INTRODUCTION

1.1 General

What is pollution? Pollution refers to the introduction of impurities into the natural environment, which can cause adverse changes. Pollution isn't always caused by chemical substances like particulates (smoke, dust); different forms of energy like sound, heat, or light can also cause pollution. These impurities that cause pollution are called pollutants. In today's era, pollution has become a cause of loss of life. According to a new study in the journal "The Lancet Planetary Health," pollution was responsible for approximately 9.0 million premature deaths in 2019.

Even in minute amounts, pollution can impact the ecological balance. Pollutants can make their way up the food chain and eventually find their way into the human body. Types of Pollution

There are different types of pollution, which are either caused by natural events or by man-made activities (like Vehicles, factories, nuclear wastes, Construction, etc) These are further classified as

- Air Pollution
- Water Pollution
- Soil Pollution
- Noise Pollution

Along with these, some other types of pollution exist such as Light pollution, thermal pollution, and radioactive pollution. The latter is much rarer than other types, but it is the deadliest.

1.1.1 Air Pollution

Air pollution is the contamination of the indoor or outdoor environment by any type of chemical, physical, or biological agent that alters the natural characteristics of the atmosphere. Household combustion devices, motor vehicles, industrial facilities and forest fires are main sources of air pollution. Pollution of major public health concern include Particulate matter, carbon monoxide, ozone, nitrogen dioxide and sulfur dioxide. Both Outdoor and indoor air pollution cause respiratory, cardiovascular diseases and other diseases and are important sources of morbidity and mortality

Air pollution refers to the release of harmful contaminants (chemicals, toxic gases, particulates, biological molecules, etc.) into the atmosphere. Common gaseous pollutants include carbon monoxide, sulfur dioxide, chlorofluorocarbons (CFCs) and nitrogen oxides produced by Motor vehicles, industry.

1.1.2 Water Pollution

Water pollution is the contamination of water sources by substances which make the waterunusable for drinking, cooking, cleaning, swimming, and other activities. Pollutants include

chemicals, trash, bacteria, and parasites. All forms of pollution eventually make their way to water. Air pollution settles onto lakes and oceans. Land pollution can seep into an underground stream, then to a river, and finally to the ocean. Thus, waste dumped in a vacant lot can eventually pollute a water supply.

1.1.2.1 What Is Water Pollution (According To WHO)

Polluted water is water that has been contaminated to the extent that it is unusable and can cause diseases that kill more than 500,000 people worldwide each year. Water pollutants can take many forms and originate from various sources, such as agricultural runoff, industrial discharges, sewage, and waste disposal. Some pollutants, such as bacteria, viruses, and parasites, can cause waterborne diseases in humans and animals, while fertilizers and pesticides can create nutrient imbalances that lead to harmful algal blooms and oxygen depletion, harming aquatic organisms. Pharmaceuticals, personal care products, nitrates, phosphates, plastics, microplastics, faecal waste, and radioactive substances can also pose hazards to aquatic ecosystems. Therefore, regular water quality testing is essential to detect the presence of pollutants and their levels, enabling authorities to take appropriate measures to prevent further contamination and ensure water sources are safe for human consumption and the environment.

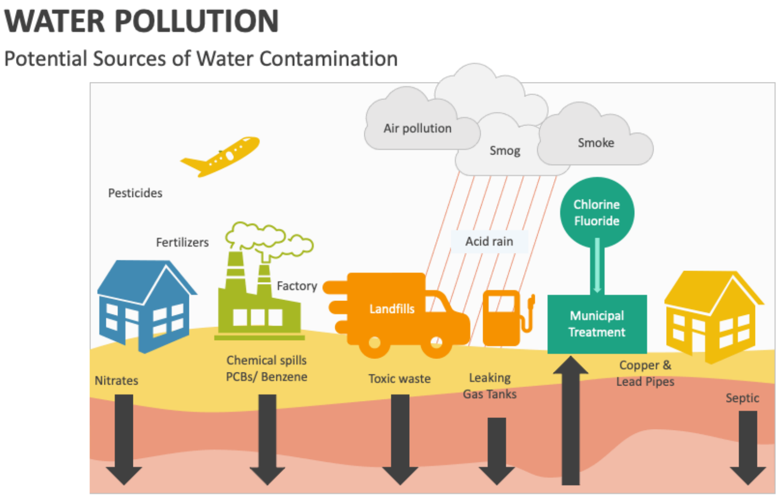


Figure1 :(water pollution)

Some of the Significant causes of water pollution include:

- → Dumping solid wastes in water bodies
- → Disposing untreated industrial wastage into water bodies
- → Connecting sewage to rivers, lakes, oceans

1.1.3 Soil Pollution

Soil pollution is the contamination of the soil with pollutants, toxic chemicals or any contaminant in such a quantity that reduces soil quality and makes it inhabitable to organisms such as insects and other microbes. Or it can be referred to as the addition of chemicals to the soil in quantities that are toxic to the environment and its residents. This addition is mostly by human activities such as mining, modern practices in agriculture, deforestation, indiscriminate dumping of human generated trash and unregulated disposalof untreated wastes of various industries

Some of the common causes of soil pollution are:

Improper industrial waste disposal

- → Oil Spills
- → Acid rain which is caused by air pollution
- → Mining activities
- → Intensive farming and agrochemicals (like fertilisers and pesticides)

→ Industrial accidents

The effects of soil pollution are numerous. Specific wastes, such as radioactive waste become particularly hazardous when they are not well-contained. A well-documented example is a nuclear accident in Chernobyl, which has left an area of 2,600 km^2 uninhabitable forseveral thousand years.

Other effects of soil pollution include:

→ Loss of soil nutrients, which renders the soil unfit for agriculture

→ Impacts the natural flora and fauna residing in the soil

→ Degrades vegetation due to the increase of salinity of the soil

→ Toxic dust (such as silica dust) can cause respiratory problems or even lung cancer

1.1.4 Noise Pollution

Noise pollution refers to the unwanted and excessive sounds in the environment that can cause harm or annoyance to humans and animals. It is a form of environmental pollution that negatively impacts physical and mental health, as well as quality of life. Common sources of noise pollution include transportation, construction sites, industrial activities, and recreational activities. Exposure to noise pollution can lead to a variety of health problems, including hearing loss, high blood pressure, sleep disturbances, stress, and anxiety. It can also negatively affect wildlife, causing physical harm or death. To reduce noise pollution, measures such as implementing noise control regulations, soundproofing buildings and homes, using quieter transportation, and reducing the use of noisy equipment can be taken. Individuals can also use noise-cancelling headphones and turn down electronic device volume to reduce their contribution to noise pollution.

1.2 Why Air Pollution

Air Pollution is now considered to the world's largest environmental health threat, accounting for7 million deaths around the world every year. Air Pollution causes and exacerbates a number of diseases, ranging from asthma to cancer, pulmonary illnesses and heart disease. Outdoor air pollution and particulate matter, one of its major components, have been classified as carcinogenic to humans by the international Agency for Research on cancer.

In accordance with recent estimates by the World Health Organization, exposure to air pollution is thus a more important risk factor for major non-communicable diseases than previously thought. Air pollution is the largest contributor to the burden of disease from the environment.

The main substances affecting health are Nitrogen oxides (NOx), sulphur oxides (Sox),ozone andparticulate matter with the latter – especially particulate matter below 2.5 microns (PM 2.5) being of greatest concern, as these tiny particles penetrate deep into the lungs, affecting both the respiratory and vascular systems. Both extent and duration of the exposure influence health outcomes.

Polluted air is creating a health emergency

There is no doubt today that air Pollution is a global public health emergency. It threatens everyone from unborn babies to children walking to school, to women cooking over open fires.

1.2.1 Effects/Impact of Air Pollution on Health

According to the WHO, air pollution is linked to approximately seven million premature deaths annually due to exposure from both outdoor and household air pollution. This makes it the world's largest environmental health risk, comparable with ―traditional‖ health risks such as smoking, high cholesterol, high blood sugar and obesity. Figure below shows that most air pollution-related deaths are from heart disease and stroke, followed by chronic obstructive pulmonary disease, acute and chronic respiratory conditions and cancers.

https://www.collidu.com/media/catalog/product/img/a/1/

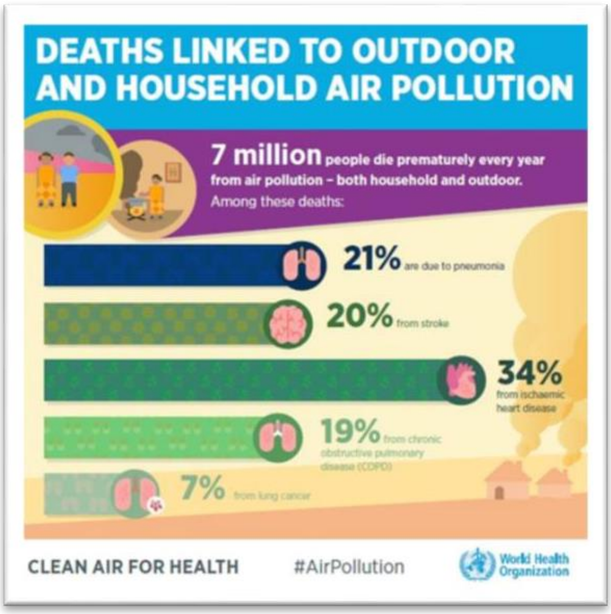


*Figure 2 :**Deaths Linked To Outdoor And Household Air** Pollution (World Health Organization -Clean Air For Health)*

The above image shows that ambient (outdoor) air pollution alone caused some 4.2 million deathin 2016 due to exposure to small particulate matter of 2.5 micron or less in diameter ($PM_{2.5}$), which cause cardiovascular and respiratory disease, and cancers. On the other hand, household (indoor) air pollution from cooking with polluting fuels and technologies caused an estimated 3.8million deaths in same period.

Deaths due to House hold Air Pollution 2019

Figure 3: Household Air Pollution (World Health Organization -Clean Air For Health)

Figure 4: Air Pollution-The Silent Killer(World Health Organization-Clean Air For Health-2019)

1.2.1.1 Children are most at risk:

Children are particularly vulnerable to the damaging health effects of air pollution due to their unique susceptibility and exposure: Children's respiratory tracts are more permeable, their breathing rate is twice as much as adults, and they take in more air per kilogram (kg) of their body weight. Children's bodies, especially their lungs and brains, are still developing, with narrower blood vessels. And their immune systems are weaker than adults; hence, polluted air affects children more than adults.
Air pollution contributed to nearly 500,000 deaths among infants in 2019, with most related to complications of low birth weight and preterm birth. Close to half of pneumonia deaths in children under 5 are caused by particulate matter from household air pollution. 20% of newborn deaths are attributed to air pollution, and women in countries with low sociodemographic development are at risk for adverse birth outcomes. Air pollution affects a child's development, learning, and well-being throughout their lifetime. These statistics are from the State of Global Air 2020 estimates and the World Health Organization, highlighting the significant global burden of diseases caused by air pollution.

Figure 5: Impact Of Air Pollution On Children's Health (World Health Organization)

1.2.2 World worst air pollution disasters

Air pollution disasters have occurred around the world, with severe impacts on the environment and public health. Here are some of the worst air pollution disasters in history
1)Bhopal Gas Tragedy (India, 1984) – A gas leak at a pesticide plant in Bhopal, India, exposed

***1.2.2.1 Bhopal disaster**,*

chemical leak in 1984 in the city of Bhopal, Madhya Pradesh state, India. At the time, it was called the worst industrial accident in history.
On December 3, 1984, about 45 tons of the dangerous gas methyl isocyanate escaped from an insecticide plant that was owned by the Indian subsidiary of the American firm Union Carbide Corporation. The gas drifted over the densely populated neighbourhoods around the plant, killing thousands of people immediately and creating a panic as tens of thousands of others attempted to flee Bhopal. The final death toll was estimated to be between 15,000 and 20,000. Some half a million survivors suffered respiratory problems, eye irritation or blindness, and other maladies resulting from exposure to the toxic gas; many were awarded compensation of a few hundred dollars. Investigations later established that substandard operating and safety

procedures at the understaffed plant had led to the catastrophe. In 1998 the former factory site was turned over to the state of Madhya Pradesh.

1.2.2.2 Great Smog of London

The Great Smog of London was a severe air pollution event that occurred in London, England, in December 1952. The smog was caused by a combination of industrial pollution, emissions from coal fires, and weather conditions that trapped the pollutants close to the ground.
The smog lasted for five days, from December 5 to December 9, 1952, and was responsible for an estimated 4,000 to 12,000 deaths. The exact number of deaths is difficult to determine because many of the deaths were due to respiratory and cardiovascular illnesses that were exacerbated by the smog, rather than being directly caused by it.
During the smog, visibility was severely reduced, with some areas of the city experiencing visibility of less than 1 meter. The smog also caused widespread disruption, with transport systems and businesses forced to shut down.
The Great Smog of London was a significant event in the history of air pollution control. It led to the introduction of the Clean Air Act in 1956, which established smoke-free zones and regulated the burning of fuels in urban areas. The event also served as a warning of the dangers of air pollution and the need for environmental regulation to protect public health
environmental disaster, England, United Kingdom [1952]

.

1.2.2.3 *Chernobyl Disaster (Ukraine, 1986)*

The Chernobyl disaster is one of the most catastrophic nuclear power plant accidents in history. It occurred on April 26, 1986, at the Chernobyl Nuclear Power Plant in Pripyat, Ukraine, then part of the Soviet Union.
The disaster was caused by a combination of human error, design flaws in the reactor, and a lack of proper safety protocols. During a routine test, a power surge occurred, causing an explosion and a fire that released large amounts of radioactive materials into the atmosphere.
The explosion and subsequent release of radiation had devastating effects on the environment and public health. The disaster resulted in immediate deaths of two plant workers, and long-term health effects on hundreds of thousands of people, including increased rates of cancer, birth defects, and other health issues.
The effects of the Chernobyl disaster also had a significant impact on the environment, with large areas of land surrounding the plant remaining contaminated with radiation for decades after the accident.
The response to the disaster was a massive effort that involved the evacuation of thousands of people from the surrounding areas, the construction of a containment structure around the damaged reactor, and the implementation of measures to mitigate the environmental and health impacts of the disaster.
The Chernobyl disaster serves as a reminder of the importance of proper safety protocols and regulations in the nuclear industry and the need for ongoing monitoring and management of nuclear waste and facilities.

1.3 Air Pollution – Introduction

Air pollution is a mix of particles and gases that can reach harmful concentrations both outside and indoors. Its effects can range from higher disease risks to rising temperatures. Soot, smoke, mold, pollen, methane, and carbon dioxide are a just few examples of common pollutants
The following are **primary pollutants**, which means they are emitted directly into the atmosphere by a particular source.

- Volatile organic compounds (VOCs) such as gases from evaporating gasoline.
- Carbon monoxide (CO), which is an odorless gas that is produced by combustion.
- NOx, which includes nitric oxide (NO) and nitrogen dioxide (NO_2).
- Sulfur dioxide (SO_2), which is produced through the combustion of coal.
- Particulate matter (PM), which are small suspended particles.
- Lead, which we used to add to our gasoline.

Primary pollutants can combine with other chemicals in the atmosphere and produce **secondary pollutants**

- NOx can produce nitric acid (HNO_3).
- Sulfur dioxide (SO_2) can produce sulfuric acid (H_2SO_2).
- Ozone (O_3), which is produced through chemical reactions between nitrogen dioxide, VOCs, and sunlight.

.

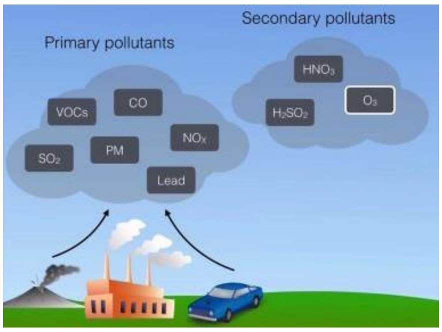


Figure 6: AIR Pollutants ***(Majid Farooq et al.)***

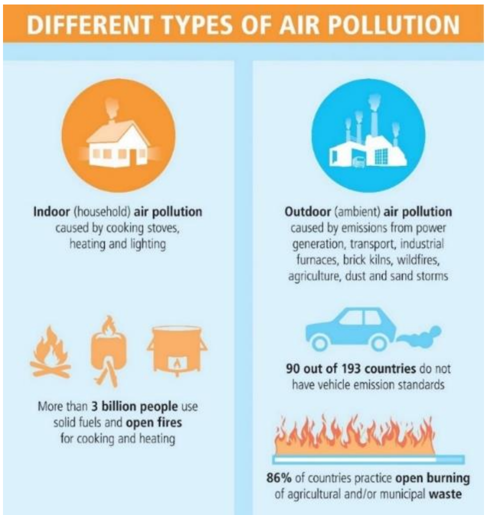


Figure 7: Different Types Of Air Pollution (Majid Farooq et al)

1.3.1 Household(Indoor)Air pollution

Indoor air is air within a building such as your home, office, classroom, hospital, shopping centeror gym etc. We say it as „Indoor Air Pollution" if indoor air is contaminated by smoke, chemicals, smells or particles

Indoor air pollution involves exposures to particulates, carbon oxides, and other pollutants carried by indoor air or dust. Examples include:
Gases (carbon monoxide, radon, etc.)
Household products and chemicals

- Building materials (asbestos, formaldehyde, lead, etc.)
- Indoor allergens (cockroach and mouse dropping, etc.)
- Tobacco smoke

1.3.1.1 Key facts given by WHO on Household air pollution and health

Around 2.4 billion people worldwide(around a third of the global population) cook usingopen fires or inefficient stoves fuelled by kerosene, biomass (wood, animal dung and crop waste) and coal. Such inefficient cooking and heating practices produce high levelsof household (indoor) air pollution which includes a range of health damaging pollutantssuch as fine particles and carbon monoxide.

→ Household air pollution was responsible for an estimated 3.2 million deaths per year in 2020, including over 237000 deaths of children under age of 5
→ Household air pollution exposure leads to noncommunicable diseases including stroke, ischemic heart disease, chronic obstructive pulmonary disease (COPD) and lung cancer
→ Women and children, typically responsible for household chores such as cooking collecting firewood, bear the greatest health burden from the use of polluting fuels and technologies in homes.
→ It is essential to expand use of clean fuels and technologies to reduce household air pollution and protect health. These include solar, electricity, biogas, liquefied petroleum gas (LPG), natural gas, alcohol fuels, as well as biomass stoves that meet the emission targets in the WHO Guidelines.

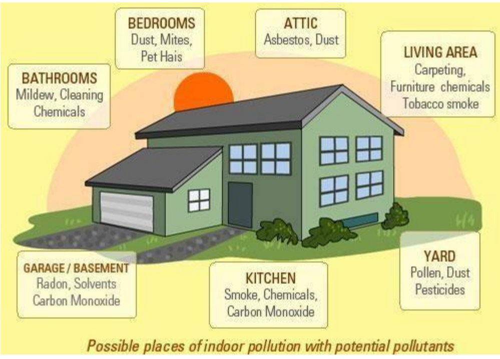


Figure 8: possible places of indoor air pollution(INTOSAI WGEA Work Plan 2017-2019)

1.3.2 Outdoor Air Pollution (Ambient Air Pollution)

Outdoor air pollution, usually caused by emissions from transportation, power generations, agriculture, and open burning. The chemicals in Ambient air pollution is different from that of indoor air pollution
The main point is that when pollution occurs in the air, it can easily travel and spread, and because we breathe in the air, we cannot easily avoid it

- Fine particles produced by the burning of fossil fuels (i.e. the coal and petroleum used in trafficand energy production)
- Noxious gases (sulphur dioxide, nitrogen oxides, carbon monoxide, chemical vapors, etc.)
- Ground-level ozone (a reactive form of oxygen and a primary component of urban

smog)

- Tobacco smoke

Power plants, factories, and vehicles spew out harmful gases and small particles that can penetrate deep into all people's lungs, particularly the children who are the vulnerable social group. In a broad day light, ground level ozone may be formed by the reaction of Nitric Oxides (NOx) and Volatile Organic Compounds (VOCs) under the influence of sunlight, which can trigger asthma attacks. Air pollution does not respect national borders.

Heavy metals and persistent organic pollutants are carried by winds, contaminating water and soil far from their origin. In the 1990s, forest fires mainly from Indonesia, caused a haze of smoke to hang for months over neighboring Southeast Asian countries

Figure 9: Sources Of Air Pollution (World Health Organization)

1.3.2.1 Key Facts by WHO on Ambient (outdoor) Air Pollution

- Air pollution is one of the greatest environmental risk to health. By reducing air pollutionlevels, countries can reduce the burden of disease from stroke, heart disease, lung cancer, and both chronic and acute respiratory diseases, including asthma.
- The lower the levels of air pollution, the better the cardiovascular and respiratory healthof the population will be, both long- and short-term.
- The WHO Air Quality Guidelines: Global Update 2021 provide an assessment of health effects of air pollution and thresholds for health–harmful pollution levels.
- In 2019, 99% of the world population was living in places where the WHO air quality guidelines levels were not met.
- Ambient (outdoor air pollution) in both cities and rural areas was estimated to cause 4.2 million premature deaths worldwide in 2016.
- Some 91% of those premature deaths occurred in low- and middle-income countries, and the greatest number in the WHO South-East Asia and Western Pacific regions.
- Policies and investments supporting cleaner transport, energy-efficient homes, power generation, industry and better municipal waste management would reduce key sources of outdoor air pollution.
- In addition to outdoor air pollution, indoor smoke is a serious health risk for some 2.4 billion people who cook and heat their homes with biomass, kerosene fuels and coal**.**

1.3.2.2 Main Causes Of Anthropogenic Air Pollution

- How much pollution we breathe in is dependent on many factors, such as access to clean energy for cooking and heating, the time of day and the weather. Rush hour is an obvious source of local pollution, but air pollution can travel long distances, sometimes across continents on international weather patterns.

- Nobody is safe from this pollution, which comes from five main human sources. These sources spew out a range of substances including carbon monoxide, carbon dioxide, nitrogen dioxide, nitrogen oxide, ground-level ozone, particulate matter, sulfur dioxide, hydrocarbons, and lead–all of which are harmful to human health.

1.3.2.2.1 Agriculture

There are two major sources of air pollution from agriculture: livestock, which produces methane and ammonia, and the burning of agricultural waste. Methane emissions contribute to ground-level ozone, which causes asthma and other respiratoryillnesses. Methane is also a more potent global warming gas than carbon dioxide – its impact is 34 times greater over a 100- year period. Around 24 percent of all greenhouse gases emitted worldwide come agriculture, forestry and other land-use.
There are many ways to reduce air pollution from this source. People can move to a plant-based diet and/or reduce food waste,while farmers can reduce methane from livestock by optimizing feed digestibility and improving grazing and grassland management.

1.3.2.2.2 Household

The main source of household air pollution is the indoor burning of fossil fuels, wood and other biomass-based fuels to cook,heat and light homes. Around 3.8 million premature deaths are caused by indoor air pollution each year, the vast majority of them in the developing world.
Out of 193 countries, 97 countries have increased the percentage of households that have access to cleaner burning fuels to over 85 percent. However, 3 billion people continue to use solid fuels and open fires for cooking, heating, and lighting. The adoption of cleaner, more modern stoves and fuels can reduce the risks of illness and save lives.

1.3.2.2.3 Industry

In many countries, energy production is a leading source of air pollution. Coal-burning power plants are a major contributor, while diesel generators are a growing concern in off-grid areas. Industrial processes and solvent use, in the chemical and miningindustries, also pollute the air.
Policies and programmes aimed at increasing energy efficiency and production from renewable sources have a direct impacton a country's air quality. At the moment, 82 countries out of 193 have incentives that promote investment in renewable energyproduction, cleaner production, energy efficiency and pollution control.

1.3.2.2.4 Transport

The global transport sector accounts for almost one-quarter of energy-related carbon dioxide emissions and this proportion is rising. Transport emissions have been linked to nearly 400,000 premature deaths. Almost half of all deaths by air pollution from transport are caused by diesel emissions, while those living closest to major traffic arteries are up to 12 percent more likely to be diagnosed with dementia.

Reducing vehicle emissions is an important intervention to improve air quality, especially in urban areas. Policies and standards that require the use of cleaner fuels and advanced vehicle emissions standards can reduce vehicle emissions by 90 percent or more.

1.3.2.2.5 Waste

Open waste burning and organic waste in landfills release harmful dioxins, furans, methane, and black carbon into the atmosphere. Globally, an estimated 40 percent of waste is openly burned. The problem is most severe in urbanizing regions and developing countries. Open burning of agricultural and municipal waste is practiced in 166 out of 193 countries.
Improving the collection, separation, and disposal of solid waste reduces the amount of waste that is burned or landfilled. Separating organic waste and turning it into compost or bioenergy improves soil fertility and provides an alternative energy source. Reducing the estimated one-third of all food that is lost or wasted can also improve air quality.

1.3.2.2.6 Other Sources

Not all air pollution comes from human activity. Volcanic eruptions, dust storms and other natural processes also cause problems. Sand and dust storms are particularly concerning. Fine particles of dust can travel thousands of miles on the back ofthese storms, which may also carry pathogens and harmful substances, causing acute and chronic respiratory problems.

1.4 Air pollution pathway

Exposure to air pollution is a risk factor that causes health impacts (Smith 1993, McGranahan and Murray 2003, U.S. Environmental Protection Agency 2009a). Epidemiological risk is the probability that a disease, injury, or infection will occur. The risk assessment of air pollution follows the air pollution pathway (Figure 1) from sources through emissions, concentrations, exposures, doses, to health impacts (Smith 1993, McGranahan and Murray 2003, U.S. Environmental Protection Agency 2009a). Sources are the origin of the pollutant, generally the quantity and quality of fuel used. Emissions are air pollutants released from the source and are characterized by the environment, transported, and transformed. Concentrations are the amount of an air pollutant in space and time. Exposures are concentrations of air pollutants that are breathed in and depend on pathways, durations, intensities, and frequencies of contact with the pollutant. Doses are how much of the exposure is deposited in the body. Health impacts accrue from doses, can be acute (short-term) or chronic (long-term), and are non-specific in that they have many risk factors. Monitoring and intervention can occur at any stage along this pathway. Health impacts are the primary risk indicators, though control measures at this stage are often too late and complicated due to their non-specific nature. Doses are also too late in the air pollution pathway and are poorly understood for many pollutants. Control measures and standards generally focus on sources, emissions, and concentrations, with recent efforts targeting exposures.

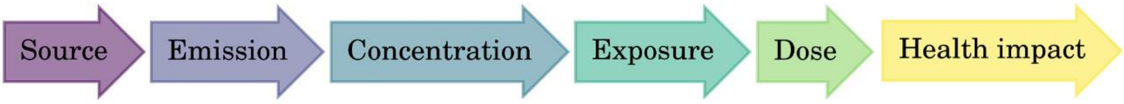


Figure 10: *Air pollution pathway (Smith 1993, McGranahan and Murray 2003, U.S. Environmental Protection Agency 2009a).*

1.4.1 Major Air Pollutants Their Sources And Effects

Air pollution is a real public health and environmental problem that can lead to-among other things-global warming, acid rain,and the deterioration of the ozone layer. This chart names some common pollutants, their sources, and their effect on the environment.

Pollutant	Sources	Effects
Ozone. A gas that can be found in two places. Near the ground (the troposphere), it is a major part of smog. The harmful ozone in the lower atmosphere should not be confused with the protective layer of ozone in the upper atmosphere (stratosphere), which screens out harmful ultraviolet rays.	Ozone is not created directly, but is formed when nitrogen oxides and volatile organic compounds mix in sunlight. That is why ozone is mostly found in the summer. Nitrogen oxides come from burning gasoline, coal, or other fossil fuels. There are many types of volatile organic compounds, and they come from sources ranging from factories to trees.	Ozone near the ground can cause a number of health problems. Ozone can lead to more frequent asthma attacks in people who have asthma and can cause sore throats, coughs, and breathing difficulty. It may even lead to premature death. Ozone can also hurt plants and crops.
Carbon monoxide. A gas that comes from the burning of fossil fuels, mostly in cars. It cannot be seen or smelled.	Carbon monoxide is released when engines burn fossil fuels. Emissions are higher when engines are not tuned properly, and when fuel is not completely burned. Cars emit a lot of the carbon monoxide found outdoors. Furnaces and heaters in the home can emit high concentrations of carbon monoxide, too, if they are not properly maintained.	Carbon monoxide makes it hard for body parts to get the oxygen they need to run correctly. Exposure to carbon monoxide makes people feel dizzy and tired and gives them headaches. In high concentrations it is fatal. Elderly people with heart disease are hospitalized more often when they are exposed to higher amounts of carbon monoxide.
Nitrogen dioxide.A reddish-brown gas that comes from the burning of fossil fuels. It has astrong smell at high levels.	Nitrogen dioxide mostly comes from power plants and cars. Nitrogen dioxide is formed in two ways-when nitrogen in the fuel is burned, or when nitrogen in the air reacts with oxygen at very high temperatures. Nitrogen dioxide can also react in the atmosphere to form ozone, acidrain, and particles.	High levels of nitrogen dioxide exposure can give people coughs and can make them feel short of breath. People who are exposed to nitrogen dioxide for a long time have a higher chance of getting respiratory infections. Nitrogendioxide reacts in the atmosphere to form acid rain, which can harm plants and animals.
Particulate matter. Solid or liquid matter that is suspended in the air. To remain in the air, particles usually must be less than 0.1-mm wideand can be as small as 0.00005 mm.	Particulate matter can be divided into two types- coarse particles and fine particles. Coarse particles are formed from sources like road dust, sea spray, and construction. Fine particles are formed when fuel is burned in automobiles and power plants.	Particulate matter that is small enough can enterthe lungs and cause health problems. Some of these problems include more frequent asthma attacks, respiratory problems, and premature death.
Sulfur dioxide. A corrosive gas that cannot be seen or smelled at low levels but can have a "rotten egg" smell at high levels.	Sulfur dioxide mostly comes from the burning of coal or oil in power plants. It also comes from factories that make chemicals, paper, or fuel. Like nitrogen dioxide, sulfur dioxide reacts in theatmosphere to form acid rain and particles.	Sulfur dioxide exposure can affect people who have asthma or emphysema by making it more difficult for them to breathe. It can also irritate people's eyes, noses, and throats. Sulfur dioxidecan harm trees and crops, damage buildings, and make it harder for people to see long distances.

Lead. A blue-gray metal that is very toxic and is found in a number of forms and locations.	Outside, lead comes from cars in areas where unleaded gasoline is not used. Lead can also come from power plants and other industrial sources. Inside, lead paint is an important source of lead, especially in houses where paint is peeling. Lead in old pipes can also be a source of lead in drinking water.	High amounts of lead can be dangerous for small children and can lead to lower IQs and kidney problems. For adults, exposure to lead can increase the chance of having heart attacksor strokes.
Toxic air pollutants. A large number of chemicals that are known or suspected to cause cancer. Some important pollutants in thiscategory include arsenic, asbestos, benzene, and dioxin.	Each toxic air pollutant comes from a slightly different source, but many are created in chemical plants or are emitted when fossil fuelsare burned. Some toxic air pollutants, like asbestos and formaldehyde, can be found in building materials and can lead to indoor air problems. Many toxic air pollutants can also enter the food and water supplies.	Toxic air pollutants can cause cancer. Some toxic air pollutants can also cause birth defects.Other effects depend on the pollutant, but can include skin and eye irritation and breathing problems.
Stratospheric ozone depleters. Chemicals that can destroy the ozone in the stratosphere.These chemicals include chlorofluorocarbons (CFCs), halons, and other compounds that include chlorine or bromine.	CFCs are used in air conditioners and refrigerators, since they work well as coolants. They can also be found in aerosol cans and fire extinguishers. Other stratospheric ozone depleters are used as solvents in industry.	If the ozone in the stratosphere is destroyed, people are exposed to more radiation from the sun (ultraviolet radiation). This can lead to skin cancer and eye problems. Higher ultraviolet radiation can also harm plants and animals.
Greenhouse gases. Gases that stay in the air for a long time and warm up the planet by trapping sunlight. This is called the “greenhouseeffect” because the gases act like the glass in a greenhouse. Some of the important greenhousegases are carbon dioxide, methane, and nitrous oxide.	Carbon dioxide is the most important greenhouse gas. It comes from the burning of fossil fuels in cars, power plants, houses, and industry. Methane is released during the processing of fossil fuels, and also comes fromnatural sources like cows and rice paddies. Nitrous oxide comes from industrial sources anddecaying plants.	The greenhouse effect can lead to changes in the climate of the planet. Some of these changes might include more temperature extremes, higher sea levels, changes in forest composition, and damage to land near the coast. Human health might be affected by diseases that are related to temperature or by damage to land and water.

***Table 1:** Major Air Pollutants Their Sources And Effects (**Jonathan Levy**)*

1.4.1.1 Atmospheric Aerosols: What Are They, And Why Are They So Important?

Aerosols are minute particles suspended in the atmosphere. When these particles are sufficiently large, we notice their presence as they scatter and absorb sunlight. Their scattering of sunlight can reduce visibility (haze) and redden sunrisesand sunsets.

Aerosols interact both directly and indirectly with the Earth's radiation budget and climate. As a direct effect, the aerosolsscatter sunlight directly back into space. As an indirect effect, aerosols in the lower atmosphere can modify the size of cloud particles, changing how the clouds reflect and absorb sunlight, thereby affecting the Earth's energy budget.

Aerosols also can act as sites for chemical reactions to take place (heterogeneous chemistry). The most significant of these reactions are those that lead to the destruction of stratospheric ozone. During winter in the polar regions, aerosols grow to form polar stratospheric clouds. The large surface areas of these cloud particles provide sites for chemical reactions to take place.

These reactions lead to the formation of large amounts of reactive chlorine and, ultimately, to thedestruction of ozone in the stratosphere. Evidence now exists that shows similar changes in stratospheric ozone concentrations occur after major volcanic eruptions, like Mt. Pinatubo in 1991, where tons of volcanic aerosols are blown into the atmosphere.

Volcanic Aerosol

Three types of aerosols significantly affect the Earth's climate. The first is the volcanic aerosol layer which forms in the stratosphere after major volcanic eruptions like Mt. Pinatubo. The dominant aerosol layer is actually formed by sulfur dioxide gas which is converted to droplets of sulfuric acid in the stratosphere over the course of a week to several monthsafter the eruption. Winds in the stratosphere spread the aerosols until they practically cover the globe. Once formed, these aerosols stay in the stratosphere for about two years. They reflect sunlight, reducing the amount of energy reaching the lower atmosphere and the Earth's surface, cooling them.

Desert Dust

The second type of aerosol that may have a significant effect on climate is desert dust. Pictures from weather satellites often reveal dust veils streaming out over the Atlantic Ocean from the deserts of North Africa. Fallout from these layers has been observed at various locations on the American continent. Similar veils of dust stream off deserts on the Asian continent. The particles in these dust plumes are minute grains of dirt blown from the desert surface. They are relatively large for atmospheric aerosols and would normally fall out of the atmosphere after a short flight if they were not blown to relatively high altitudes (15,000 ft. and higher) by intense dust storms. Because the dust is composed of minerals, the particles absorb sunlight as well as scatter it. Through absorption of sunlight, the dust particles warm the layer of the atmosphere where they reside. This warmer air is believed to inhibit the formation of storm clouds. Through the suppression of storm clouds and their consequent rain, the dust veil is believed to further desert expansion.

Human-Made Aerosol

The third type of aerosol comes from human activities. While a large fraction of human-made aerosols come in the form of smoke from burning tropical forests, the major component comes in the form of sulfate aerosols created by the burningof coal and oil. The concentration of human-made sulfate aerosols in the atmosphere has grown rapidly since the start ofthe industrial revolution. At current production levels, human-made sulfate aerosols are thought to outweigh the naturally produced sulfate aerosols. The concentration of aerosols is highest in the northern hemisphere where industrial activity iscentered. The sulfate aerosols absorb no sunlight but they reflect it, thereby reducing the amount of sunlight reaching theEarth's surface. Sulfate aerosols are believed to survive in the atmosphere for about 3-5 days. The sulfate aerosols also enter clouds where they cause the number of cloud droplets to increase but make the droplet sizes smaller. The net effectis to make the clouds reflect more sunlight than they would without the presence of the sulfate aerosols.

1.4.2 Emissions

Emissions are the release of pollutants into the air from human activities such as transportation, industrial processes, and agriculture. The main pollutants released include carbon monoxide, nitrogen oxides, sulfur dioxide, particulate matter, and volatile organic compounds. Transportation is a significant source of emissions, as are industrial processes and agriculture. Regulations and standards have been implemented to control emissions from these sources, including the use of emission control technologies and cleaner energy sources.

1.4.3 Concentration

1.4.3.1 New Who Air Quality Guidelines

- If target levels of pollution are implemented by governments, the WHO"s new air qualityguidelines will save lives by reducing preventable airborne pollutant deaths.
- Air Pollution and Health go hand in hand. The rise in Air pollution increases the associated health impacts for humans as well as the environment. This led WHO toupdate the existing Air Quality Guidelines. The guidelines were updated on 22 September, 2021.
- The updated guidelines by WHO will help policy makers to make strategy for air quality management and the aim is to save lives globally. These guidelines are updated after 16 years. New guidelines have reduced the ideal concentration of pollutants in air and aims to reduce 48 % of total global death burden due to PM2.5 solely. PM 2.5 is a knowncarcinogen according to WHO-IARC (International Agency for Research on Cancer).

1.4.3.1.1 Aim to change Air Quality Guidelines (AQGs):

- 80% reduction in global death due to PM 2.5
- Contributes in mitigating climate change
- Reduce global disease burden
- Reducing air pollutant emissions and pollution
- Protecting and ensuring health
- Reduce 57% and 60% deaths in South East Asian and African countries

POLLUTANT	AVERAGING TIME	WHO 2021 AIR QUALITY GUIDELINE	WHO 2005 AIR QUALITY GUIDELINE	CHANGE
PM2.5 (µg/m3)	Annual	5	10	-50%
	24-hours	15	25	-40%
PM10 (µg/m3)	Annual	15	20	-25%
	24-hours	45	50	-10%
O3 (µg/m3)	Peak season	60	N/A	Newly Introduced
	8-hours	100	100	Unchanged
NO2 (µg/m3)	Annual	10	40	-75%
	24-hours	25	N/A	Newly Introduced
	1-hour	200	200	Unchanged
SO2 (µg/m3)	24-hours	40	20	100%
	10-minutes	500	500	Unchanged
CO (mg/m3)	24-hours	4	N/A	Newly Introduced
	8-hours	10	N/A	Newly Introduced
	1-hour	35	N/A	Newly Introduced
	15-minutes	100	N/A	Newly Introduced

Table 2: Air Quality Guidelines (WHO)

1.4.3.2 National Ambient Air Quality Standards (NAAQS)

- The NAAQS set by the CPCB are applicable to the whole country. The CPCB draws thispower from the Air (Prevention and Control of Pollution) Act, 1981.
- These standards are essential for the development of effective management of ambient airquality.
- The first ambient air quality standards were developed in 1982 pursuant to the Air Act.
- Later, in 1994 and 1998, these standards were revised. The latest revision to the NAAQS was done in 2009 and this is the latest version being followed.
- The 2009 standards further lowered the maximum permissible limits for pollutants and made the standards uniform across the nation. Previously, industrial zones had less stringent standards as compared to residential areas.
- The compliance of the NAAQS is monitored under the National Air Quality Monitoring Programme (NAMP). NAMP is implemented by the CPCB.
- The current standards (2009) comprise 12 pollutants.

Pollutant	Time Weighted Average	Concentration in Ambient Air	
		Industrial, Residential, Rural, and Other Areas	Ecologically Sensitive Area (notified by Central Government)
Sulphur dioxide (SO_2), $\mu g/m^3$	Annual 24 hours	50 80	20 80
Nitrogen dioxide (NO_2), $\mu g/m^3$	Annual 24 hours	40 80	30 80
Particulate matter (< 10 µm) or PM_{10}, $\mu g/m^3$	Annual 24 hours	60 100	60 100
Particulate matter (< 2.5 µm) or $PM_{2.5}$, $\mu g/m^3$	Annual 24 hours	40 60	40 60
Ozone (O_3), $\mu g/m^3$	8 hours 1 hour	100 180	100 180
Lead (Pb), $\mu g/m^3$	Annual 24 hours	0.50 1.0	0.50 1.0
Carbon monoxide (CO), mg/m^3	8 hours 1 hour	02 04	02 04
Ammonia (NH_3), $\mu g/m^3$	Annual 24 hours	100 400	100 400
Benzene (C_6H_6), $\mu g/m^3$	Annual	05	05
Benzo(α)Pyrene (BaP) – particulate phase only, ng/m^3	Annual	01	01
Arsenic (As), ng/m^3	Annual	06	06
Nickel (Ni), ng/m^3	Annual	20	20

Figure 11: National Ambient Air Quality Standards -NAAQS-(CPCB)

1.4.4 Exposure& Dose

Exposure to air pollution occurs when individuals come into contact with pollutants present in the air. Exposure can occur through inhalation, ingestion, or skin contact. The level of exposure can vary depending on the duration and frequency of exposure, as well as the concentration and toxicity of the pollutants. Certain groups, such as children, elderly individuals, and those with pre-existing health conditions, may be more susceptible to the negative impacts of exposure to air pollution.

Dose refers to the amount of a pollutant that an individual is exposed to over a certain period of time. Dose is influenced by a number of factors, including the concentration of the pollutant in the air, the duration and frequency of exposure, and the individual's breathing rate. Dose-response relationships are used to evaluate the health effects of exposure to pollutants.

The impacts of exposure to air pollution can range from minor irritation to serious health effects such as respiratory and cardiovascular disease, cancer, and premature death. To minimize the negative impacts of exposure to air pollution, it is important to monitor air quality and implement policies and regulations aimed at reducing emissions and limiting exposure.

In conclusion, exposure and dose are important factors to consider when evaluating the impacts of air pollution on human health. Understanding the sources and types of pollutants, as well as the factors that influence exposure and dose, can help inform policies and regulations aimed at reducing the negative impacts of air pollution

1.4.5 Health impact

Air pollution can have significant impacts on human health, causing a wide range of respiratory, cardiovascular, and other health problems. Short-term exposure to high levels of pollutants can cause irritation of the eyes, nose, and throat, as well as headaches, dizziness, and nausea. Long-term exposure to air pollution can lead to chronic respiratory diseases such as asthma and chronic obstructive pulmonary disease (COPD), as well as cardiovascular disease, stroke, and lung cancer.

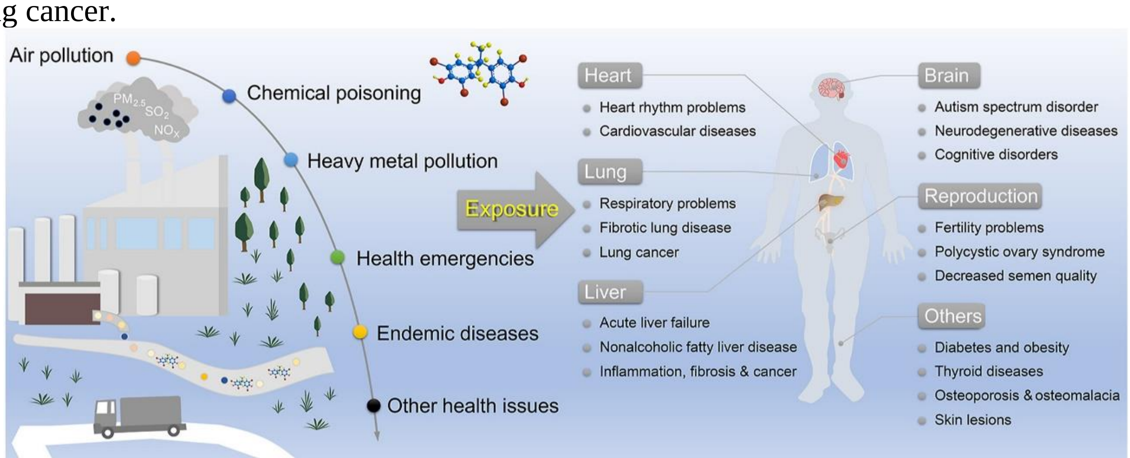


Figure 12: Air Pollution Impact (Environmental Pollution, A Hidden Culprit For Health Issues by Hanqing Xu et.al).

1.5 Government Departments Of India

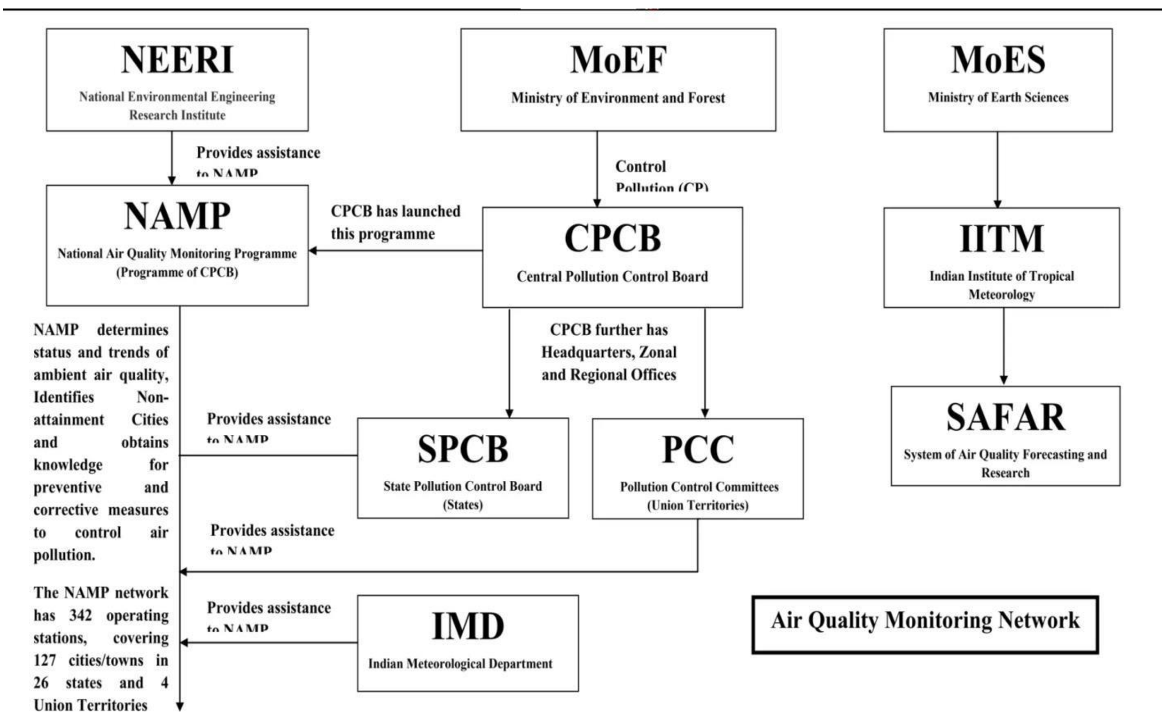


***Figure 13 :** Government Departments Of India (CPCB)*

1.5.1 Air Quality Regulations In India

Year	**Description**
1794	CPCB entrusted with the powers and functions under the air(Prevention and Control of Pollution)
1981	Central pollution control Board (CPCB)established under the water (prevention and control act)
1986	CPCB adds provisions for environment (protection) act
1994.04	National ambient air quality standards were introduced
1998.01	Environment pollution (Prevention & Control) Authority (EPCA) established to address air pollution in the national capital region (NCR)of Delhi
1998.10	National ambient air quality standards were revised
2009.11	National ambient air quality standards were revised and PM2.5 added to the list
2014.01	National air quality index (AQI) methodology was established
2016	PM2.5 is included for all manual stations under the national ambient monitoring programme (NAMP)
2016.12	Graded Response Action plan (GRAP) established to address air pollution emergencies in NCR Delhi
2018	APRIL: National Clean Air Programme (NCAP) draft released with INR 637 crores budget July: 102 non-attainment cities were announced 2019.August 20 additional non-attainment cities were announced
2018.10	EPCA reconstituted with new members from the government,academia, and civil society
2019.01	NCAP final proposal was released with INR 300 crores budget
2024	NCAP target to reduce PM2.5 pollution in the non-attainment cities by 20-30% compared to 2017 levels

Table 3:-Air Pollution in India: A Timeline of Legislations & Regulations

1.6 Definition of Air Quality Index

An air quality index is defined as an overall scheme that transforms the weighed values of individual air pollution related parameters (for example, pollutant concentrations) into a single number or set of numbers (Ott, 1978). The result is a set of rules (i.e. set of equations) that translate parameter values into a more simple form by means of numerical manipulation

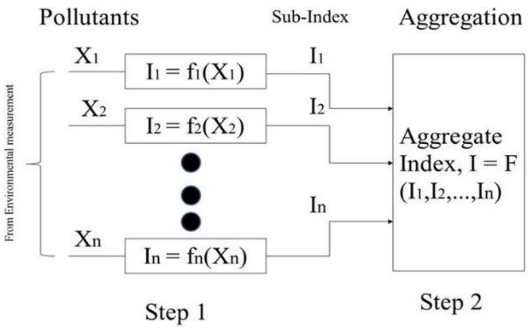


Figure 14: Formation of an Aggregated AQI (CPCB)

If actual concentrations are reported in µg/m^3 or ppm (parts per million) along with standards, then it cannot be considered as an index. At the very last step, an index in any system is to group specific concentration ranges into air quality descriptor categories.

1.6.1 Structure of an Index

Primarily two steps are involved in formulating an AQI: (i) formation of sub-indices (for each pollutant) and (ii) aggregation of sub-indices to get an overall AQI.

Formation of sub-indices (I_1, I_2,...., I_n) for n pollutant variables (X_1,X_2,......X_n) is carried out using sub-index functions that are based on air quality standards and health effects. Mathematically;

[1] $I_i=f(X_i)$, i=1, 2,.,n

Each sub-index represents a relationship between pollutant concentrations and health effects. The functional relationship between sub-indexvalue (I_i) and pollutant concentrations (X_i) is explained later in the text.

Aggregation of sub-indices, I_i is carried out with some mathematical function (described below) to obtain the overall index (I), referred to as AQI.

[2] $I=F(I_1,I_2,. \ ,I_n)$

The aggregation function usually is a summation or multiplication operation or simply a maximum operator.

Sub-indices (Step 1)

Sub-index function represents the relationship between pollutant concentration X_i and corresponding sub-index I_i. It is an attempt to reflect environmental consequences as the concentration of specific pollutant changes. It may take a variety of forms such as linear, non-linear and segmented linear. Typically, the I-X relationship is represented as follows:

[3] $I = \alpha X + \beta$

Where, α =slope of the line, β = intercept at X=0.

The general equation for the sub-index (I_i) for a given pollutant concentration (C_p); as based on 'linearsegmented principle' is calculated as:

[4] $I_i=[\{(I_{HI} - I_{LO})/(B_{HI} - B_{LO})\} * (C_p - B_{LO})] + I_{LO}$

where,

B_{HI}= Breakpoint concentration greater or equal to given concentration.

B_{LO}= Breakpoint concentration smaller or equal to given concentration.

I_{HI} =AQI value corresponding to B_{HI}

I_{LO} = AQI value corresponding to B_{LO}

Cp = Pollutant concentration

For example, we take $PM_{10}$with concentration of 85µg/m^3, $B_{HI,}$ $B_{LO,}$ $I_{HI,}$ I_{LO} values from GreaterVancouver AQI (**Table**) and using equation [4]

Sub Index (I_p) = {(100 – 50)/(100 – 50)}* (85- 50) + 50

= 85

Similarly, Sub Index can be calculated for other pollutants as well.

- → All the eight pollutants may not be monitored at all the locations. Overall AQI is calculated only if data are available for minimum three pollutants out of which one should necessarily be either PM2.5 or PM10. Else, data are considered insufficient for calculating AQI. Similarly, a minimum of 16 hours' data is considered necessary for calculating subindex.
- → All the eight pollutants may not be monitored at all the locations. Overall AQI is calculated only if data are available for minimum three pollutants out of which one should necessarily be either PM2.5 or PM10. Else, data are considered insufficient for calculating AQI. Similarly, a minimum of 16 hours' data is considered necessary for calculating subindex.
- → The sub-indices for monitored pollutants are calculated and disseminated, even if data are inadequate for determining AQI. The Individual pollutant-wise sub-index will provide air quality status for that pollutant.
- → The web-based system is designed to provide AQI on real time basis. It is an automated system that captures data from continuous monitoring stations without human intervention, and displays AQI based on running average values (e.g. AQI at 6am on a day will incorporate data from 6am on previous day to the current day).

For manual monitoring stations, an AQI calculator is developed wherein data can be fed manually to get AQI value

1.6.2 National Air Quality Index

- → Air Quality Index is a tool for effective communication of air quality status to people in terms, which are easy to understand. It transforms complex air quality data of various pollutants into a single number (index value), nomenclature and colour.
- → There are six AQI categories, namely Good, Satisfactory, Moderately polluted, Poor, Very Poor, and Severe. Each of these categories is decided based on ambient concentration values of air pollutants and their likely health impacts (known as health breakpoints). AQ sub-index and health breakpoints are evolved for eight pollutants (PM10, PM2.5, NO2, SO2, CO, O3, NH3, and Pb) for which short-term (up to 24-hours) National Ambient Air Quality Standards are prescribed.
- → Based on the measured ambient concentrations of a pollutant, sub-index is calculated, which is a linear function of concentration (e.g. the sub-index for PM2.5 will be 51 at concentration 31 µg/m3 , 100 at concentration 60 µg/m3 , and 75 at concentration of 45 µg/m3). The worst sub-index determines the overall AQI. AQI categories and health breakpoints for the eight pollutants are as follow:

❖ **BREAKPOINTS FOR AQI SCALE 0-500**

+ .AQI Category (Range)	PM_{10} 24-hr	PM2.5 24-hr	NO_2 24-hr	O_3 8-hr	CO 8-hr (mg/m^3)	SO_2 24-hr	NH_3 24-hr	Pb 24-hr
Good (0-50)	0-50	0-30	0-40	0-50	0-1.0	0-40	0-200	0-0.5
Satisfactory (51-100)	51-100	31-60	41-80	51-100	1.1-2.0	41-80	201-400	0.6 –1.0
Moderate (101-200)	101-250	61-90	81-180	101-168	2.1- 10	81-380	401-800	1.1-2.0
Poor (201-300)	251-350	91-120	181-280	169-208	10.1-17	381-800	801-1200	2.1-3.0
Very poor (301-400)	351-430	121-250	281-400	209-748*	17.1-34	801-1600	1201-1800	3.1-3.5
Severe (401-500)	430 +	250+	400+	748+*	34+	1600+	1800+	3.5+

**One hourly monitoring (for mathematical calculation only).*

Table 4: BREAKPOINTS FOR AQI SCALE 0-500 (CPCB)

- CPCB may consider reviewing the AQI breakpoints every three years after accounting the new research findings on air pollution exposure and health effects.

HEALTH STATEMENTS FOR AQI CATEGORIES

AQI	Associated Health Impacts
Good(0–50)	Minimal Impact
Satisfactory (51–100)	May cause minor breathing discomfort to sensitive people
Moderate (101–200)	May cause breathing discomfort to the people with lung disease such as asthma and discomfort to people with heart disease, children and older adults
Poor (201-300)	May cause breathing discomfort to people on prolonged exposure and discomfort to peoplewith heart disease with short exposure
Very Poor (301–400)	May cause respiratory illness to the people on prolonged exposure. Effect may be more pronounced in people with lung and heart diseases
Severe (401-500)	May cause respiratory effects even on healthy people and serious health impacts on people with lung/heart diseases. The health impacts may be experienced even during light physicalactivity

Table 5: Health Statements for AQI Categories (CPCB)

1.7 AIM:- Study Of Air Quality In State Of Bihar

The aim of your research paper could be to assess the air quality in the state of Bihar, with a focus on the city of Gaya, by examining the level of awareness of air pollution among people, determining the effect of air pollution on human health, identifying the age groups and working fields that are most affected by air pollution, determining the most significant diseases caused by air pollution, and gaining insight into the perceptions and opinions that people hold about the city of Gaya. Additionally, your research aims to forecast the PM2.5 concentration for the years 2022 and 2023 in a polluted region of Bihar using trend method, smoothing method, and model fitting. By achieving these objectives, the research aims to provide a comprehensive understanding of air quality in Bihar and identify potential strategies for improving air quality in the region.

1.8 Objectives:-

1.8.1 Secondary objective

→ To Identify The 5 most polluted cities in Bihar
→ Forecasting Pm2.5 concentration for the year 2022 & 2023 in a polluted region of Bihar by using
- Trend method
- Smoothing method
- Model fitting

1.8.2 Primary objective

→ To Examine Awareness Of Air Pollution Among People In Gaya City.
→ To Determine Effect Of Air Pollution On Human Health.
→ To Determine Which Age Group Is Mostly Affected By Air Pollution.
→ To Determine Which Working Field Is Mostly Affected By Air Pollution.
→ To Determine Which Disease Is Mostly Significant Among People In Gaya Due To Air Pollution.
→ To Gain Insight Into the Perceptions And Opinions That People Hold About The City Of Gaya.

CHAPTER 2: LITERATURE REVIEW

This chapter consists of comprehensive literature review on air pollution. The literature is grouped under the following categories.

1. Studies on air pollution.
2. Studies on the impact of air pollution on human health.
3. Studies on assessment of ambient air quality.
4. "Time Series Analysis for PM2.5 Air Quality Forecasting: A Review of Studies"

2.1 Studies On Air Pollution

Lewtas (2007) described the combustion emissions amounting to more than half the fine particle (PM2.5) air pollution and particulate organic matter. Exposure to these emissions, mutagenic and carcinogenic components such as polycyclic aromatic compounds (PAC), nitro-PAC was stated to have been extensively studied. The sources of these emissions are gasoline and diesel vehicles, burning of coal, oil, bio-mass, tobacco smoke, etc. Both short-and long-term exposure to emission was studied. The author further noted that the information revealed a high risk factor for causing lung cancer, 27 cardiopulmonary mortality and mortality due to gradual increase in exposures to air pollution apart from low birth weights of new born babies. [ref]

Lu et al. (2013) studied the hourly and daily trends of PM2.5 concentration and summarized the spatial change in PM2.5 concentration on the locations of the concentration hot spots based on data of PM2.5 concentration, wind speed, and wind direction collected at the air quality monitoring stations in 2010 at Kaohsiung city, Taiwan. The author reported the results from the correlation analysis of PM2.5 concentration and suggested that for short-term (1-hr), PM2.5 concentration could be easily affected by pollution sources around the monitoring stations, due to the atmospheric dispersion. It was recommended that there was need to further assess the influences of these high PM2.5 concentration hotspots on the local people and formulate effective strategies for pollution emission control.
[ref]

The two papers demonstrate that PM2.5 air pollution is a significant public health concern, with various sources and impacts. Lewtas (2007) emphasizes the health risks associated with exposure to PM2.5 emissions, such as lung cancer, cardiopulmonary mortality, and low birth weights. Lu et al. (2013) highlights the need to understand the spatial and temporal patterns of PM2.5 concentration in order to develop effective strategies for pollution emission control. Overall, both papers emphasize the importance of addressing PM2.5 air pollution to protect public health.

2.2 Health Effects Due To Air Pollution

Vanos et al. (2014) stated that unconditional weather and poor air quality have a bad influence on human health. The paper investigated the risk of mortality from respiratory, cardiovascular diseases on exposure to the air pollutants based on the weather in 10 major Canadian cities from 1981 to 1999. Analyses were made city-wise using Poisson generalized linear models for each

of the six weathers. This study gave an idea to the environmentalists to adapt strategies in reducing all types of pollutants in the atmosphere.
[ref]

Jayanthi and Krishnamoorthi, (2006) have analysed the seasonal variation in distribution of SO2, NOx, SPM and PM10 emissions at four different areas of Chennai, Tamilnadu South India. The authors have recorded a maximum of 1091 µg/m 3 of SPM during 2006 winter at T.Nagar and around 265 µg/m 3 of SPM at Adaiyar where as the PM10 concentration was found to be the maximum during summer season. The reason highlighted was the increasing vehicular population.
[Ref]

The Lancet Planetary health (2020) Health and economic impact of air pollution in the states of India: the Global Burden of Disease Study 2019 The high burden of death and disease due to air pollution and its associated substantial adverse economic impact from loss of output could impede India's aspiration to be a $5 trillion economy by 2024. Successful reduction of air pollution in India through state-specific strategies would lead to substantial benefits for both the health of the population and the economy.
[Ref]

Nachman and Parker (2012)evaluated the relationship between annual average ambient fine particulate matter (PM2.5) concentrations and respiratory outcomes for adults using modelled air pollution and health outcome data and to examine PM2.5 sensitivity across race ethnicity. Logistic regression was employed to investigate increases in ambient PM2.5 concentrations and self-reported prevalence of respiratory outcomes including asthma, sinusitis and chronic bronchitis.
[Ref]

The four papers demonstrate the harmful effects of air pollution on human health and the environment. They provide evidence that exposure to pollutants such as PM2.5, NOx, SPM, SO2, and others can lead to respiratory and cardiovascular diseases, premature deaths, and economic losses. The papers also suggest that pollution sources such as transportation, industrial emissions, and agricultural practices need to be effectively controlled to reduce the levels of pollutants in the atmosphere. Furthermore, the studies emphasize the importance of state-specific strategies and policies to tackle air pollution and its adverse impacts.

2.3 Studies on assessment of ambient air quality

Introduction:

Air pollution is a global environmental problem and the assessment of ambient air quality is critical for the identification and management of pollutants. Ambient air quality is affected by various factors, including industrial emissions, traffic, and biomass burning. In recent years, several studies have been conducted to assess the ambient air quality in different parts of the world. This literature review aims to provide an overview of some of the studies on the assessment of ambient air quality, their methods, and findings.

Literature Review:

In a study conducted by Chaloulakou et al. (2013), the ambient air quality in Athens, Greece, was assessed using a network of air quality monitoring stations. The study found that the levels of particulate matter (PM10) and nitrogen dioxide (NO2) exceeded the EU limit values, and the most significant sources of pollution were traffic emissions and biomass burning.

Wang et al. (2019) conducted a study in Beijing, China, to assess the impact of different pollution sources on the ambient air quality. The study used a combination of satellite remote sensing, ground-level monitoring, and air quality modeling. The results showed that traffic emissions were the primary source of PM2.5 pollution in Beijing.

In India, several studies have been conducted to assess the ambient air quality in different cities. For example, Sharma et al. (2020) assessed the air quality in Delhi, India, during the COVID-19 lockdown period. The study found a significant reduction in the levels of PM2.5, NO2, and sulfur dioxide (SO2) during the lockdown period compared to the pre-lockdown period.

Guttikunda and Calori (2013) conducted a study in Hyderabad, India, to assess the air quality in the city using a combination of monitoring data, modeling, and emissions inventory analysis. The study found that the levels of PM10 and PM2.5 exceeded the Indian National Ambient Air Quality Standards, and the most significant sources of pollution were vehicular emissions, construction activities, and industrial emissions.

Conclusion:

Assessment of ambient air quality is essential for identifying sources of pollution, evaluating the effectiveness of control measures, and reducing the adverse health effects of air pollution. The studies reviewed here highlight the importance of using a combination of monitoring data, modeling, and emissions inventory analysis to assess ambient air quality. They also highlight the need for continued monitoring and assessment to ensure that air quality standards are met and maintained.

2.4 "Time Series Analysis for PM2.5 Air Quality Forecasting: A Review of Studies"

"Time series analysis of air pollution and health impacts assessment using PM2.5 and PM10 data in Delhi, India" by S. Kumari, S. K. Goyal, and S. K. Goyal (2019).
This study analyzed the impact of air pollution on public health in Delhi, India, using PM2.5 and PM10 data. The authors used time series analysis to investigate the temporal patterns of air pollution and their effects on human health. They found that PM2.5 and PM10 levels were positively associated with hospital admissions for respiratory and cardiovascular diseases. The study also revealed that exposure to high levels of air pollution was associated with a higher risk of mortality in the population.
(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6742141/)

"Forecasting PM2.5 concentration using time series modeling in a polluted region of India" by M. K. Gupta, M. Kumar, and D. Singh (2019).

This study aimed to forecast PM2.5 concentration in a polluted region of India using time series modeling. The authors used autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) models to predict PM2.5 levels. They found that the SARIMA model performed better than the ARIMA model in forecasting PM2.5 concentrations. The study also revealed that meteorological factors, such as temperature and wind speed, had a significant impact on air quality in the region.
(https://link.springer.com/article/10.1007/s10661-019-7479-9)

"A time series model for prediction of PM2.5 concentration in Delhi-NCR, India" by D. N. D. Soni, N. Pandey, and A. Sharma (2020).

This study developed a time series model to predict PM2.5 concentration in Delhi-NCR, India. The authors used ARIMA, SARIMA, and seasonal decomposition of time series (STL) models to forecast PM2.5 levels. They found that the SARIMA model performed best in predicting PM2.5 concentrations. The study also revealed that traffic, industrial emissions, and meteorological factors were significant contributors to air pollution in the region. (https://www.sciencedirect.com/science/article/pii/S1352231020304267)

Overall, these studies demonstrate the usefulness of time series analysis in predicting PM2.5 levels and assessing the impact of air pollution on public health in India. The findings suggest that targeted interventions and policies are needed to address the sources of air pollution and protect the population from its negative effects.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Framework

3.1.1 Why India ?

3.1.1.1 IQ Air :-

- According to **IQ Air** from 2018 to 2021 India is among the top 5 most polluted countries in world based on the $PM_{2.5}$(μg/m^3).
- **Some of the facts related to top most polluted countries are given below :**
- India the third-most polluted country in the world, with an average $PM_{2.5}$ concentration of 51.90. Of the 30 most polluted cities in the world, 21 of them are located in India. The most polluted city in India and the world is Kanpur, where the city's medical college receives about 600 respiratory illness patients per month.
- India's unhealthy pollution levels are from sources such as vehicles, the burning of coal and wood, dust storms, and forest fires.
- Delhi, India's capital region, is notorious for some of the worst air in India, forcing flight cancellations, causing traffic accidents, closing of schools, and even turning the white marble walls of the Taj Mahal yellow and green.
- Rural areas are also heavily affected by pollution in India, as people rely on fuels such as wood and dung for cooking and heating and still practice the burning of crop stubble.[1]

IQ Air Data[2018-2021] :-

Rank	Country	2021	2020	2019	2018
1	Bangladesh	76.9	77.1	83.3	97.1
2	Chad	75.9	-	-	-
3	Pakistan	66.8	59	65.8	74.3
4	Tajikistan	59.4	30.9	-	-
5	India	58.1	51.9	58.1	72.5

Table 6:Top Air polluted countries 2018-2021(IQAIR)

- While the IQ Air list is respectable, it is not the only available source of air pollution data. Using data from Seattle, Washington's Institute for Health Metrics and Evaluation, the Health Effects Institute has released its own list ranking the levels of air pollution of 196 nations via the State of Global Air Report.
- India’s annual average PM2.5 levels reached 58.1 μg/m³ in 2021, ending a three-year trend of improving air quality. India's annual PM2.5 averages have now returned to pre-quarantine concentrations measured in 2019.
- India was home to 11 of the 15 most polluted cities in Central and South Asia in 2021. Delhi saw a 14.6% increase in PM2.5 concentrations in 2021 with levels rising to 96.4 μg/m³ from 84 μg/m³ in 2020. No cities in India met the WHO air quality guideline of 5 μg/m³. In 2021, 48% of India's cities exceeded 50 μg/m³, or more than 10 times the WHO guideline.

RECENT 2022 REPORT

S.No	Country	Average PM2.5 Concentation (µg/m³)
1	CHAD	89.7
2	IRAQ	80.1
3	PAKISTAN	70.9
4	BAHRAIN	66.6
5	BANGLADESH	65.8
6	BURKINA FASO	63.0
7	Kuwait	55.8
8	India	53.3

Table 7: Top Air Polluted Countries 2022 (IQAIR)

3.1.1.2 State of Global Air :-

Top 10 Countries With The Worst Air Pollution – $PM_{2.5}$ Exposure (µg/M^3)-State Of Global Air 2020 Data.

1. India - 83.2
2. Nepal - 83.1
3. Niger - 80.1
4. Qatar – 76
5. Nigeria - 70.4
6. Egypt - 67.9
7. Cameroon - 64.5
8. Bangladesh - 63.4
9. Pakistan - 62.6

The cleanest countries in the world, as determined by the Environmental Performance Index (EPI), have high air quality, clean water, and strong environmentally-friendly policies and initiatives. These countries include Switzerland, France, and Denmark.[2]

3.1.1.3 WHO Data (PM2.5)

WHO :-

- According to WHO data India is in Top 15 countries in world in Air pollution based on Concentrations of fine particulate matter (PM2.5)

SR.No.	Country	2016	Country	2017	Country	2018	Country	2019
1	Kuwait	64.37 [55.89-72.90]	Kuwait	65.63 [56.91-74.54]	Afghanistan	70.25 [50.62-97.15]	Kuwait	64.08 [55.65-72.49]
2	Afghanistan	64.00 [46.01-87.23]	Qatar	63.11 [48.58-75.64]	Kuwait	66.05 [57.26-75.10]	Egypt	63.16 [40.38-92.33]
3	Tajikistan	62.73 [42.52-90.41]	Egypt	62.05 [39.89-90.79]	Qatar	62.66 [48.36-75.02]	Afghanistan	62.49 [45.04-86.46]
4	Egypt	61.97 [39.85-90.78]	Afghanistan	61.78 [44.81-85.24]	Egypt	62.20 [39.96-90.86]	Qatar	59.04 [45.74-70.20]
5	Qatar	61.30 [47.49-73.16]	Tajikistan	59.39 [40.92-86.57]	Tajikistan	58.58 [38.93-82.46]	Saudi Arabia	57.16 [40.09-76.04]
6	Cameroon	59.20 [39.44-85.74]	Cameroon	59.30 [39.59-85.72]	Saudi Arabia	58.30 [41.15-77.33]	Cameroon	56.37 [39.06-79.81]
7	Nigeria	57.17 [37.89-79.67]	Saudi Arabia	58.47 [41.37-77.10]	Cameroon	57.80 [39.37-83.20]	Nigeria	55.64 [37.64-76.88]
8	Saudi Arabia	55.76 [39.29-73.51]	Nigeria	57.45 [38.08-80.09]	Nigeria	56.52 [37.92-78.35]	Tajikistan	53.65 [38.18-76.75]
9	Pakistan	53.65 [43.88-65.45]	Bahrain	54.72 [48.32-60.98]	Pakistan	56.50 [46.53-68.69]	Bahrain	51.82 [45.88-57.66]
10	**India**	53.00 [50.57-55.28]	Niger	52.41 [22.14-106.83]	**India**	55.48 [53.03-58.02]	**India**	50.17 [47.87-52.43]
11	Bahrain	52.90 [46.87-59.15]	Pakistan	51.98 [42.55-62.74]	Bahrain	54.49 [48.17-60.76]	Niger	50.15 [21.92-99.90]
12	Niger	52.51	**India**	51.30	Niger	50.74	Pakistan	50.13

Table 8: Top 15 Air polluted countries in world 2016-2019 (WHO)

As from given table we can see India from 2016 to 2019 ranked accordingly 10^{th},12^{th},10^{th} & 10^{th} .[3]

3.1.2 Why Bihar?

3.1.2.1 WHO

The WHO data shows that among the top 15 polluted cities in the world 11 cities were from India which includes Kanpur, Faridabad , Gaya, Varanasi, Patna, Delhi , Lucknow ,Etc.[4]

				PM10		PM2.5		
Region	iso3	Country	City/Town	Annual mean, ug/m3	note on converted PM10	Annual mean, ug/m3	Temporal coverage	note on converted PM2.5
South-East Asia (LMIC)	IND	India	Kanpur	(319)-converted value	Converted	173	>75%	Measured
South-East Asia (LMIC)	IND	India	Faridabad	(316)-converted value	Converted	172	>75%	Measured
South-East Asia (LMIC)	IND	India	Gaya	(275)-converted value	Converted	149	50% -< 75%	Measured
South-East Asia (LMIC)	IND	India	Varanasi	260	Measured	146	>75%	Measured
South-East Asia (LMIC)	IND	India	Patna	(266)-converted value	Converted	144	>75%	Measured
South-East Asia (LMIC)	IND	India	Delhi	292	Measured	143	>75%	Measured
South-East Asia (LMIC)	IND	India	Lucknow	(255)-converted value	Converted	138	>75%	Measured
Africa (Sub-Saharan) (LMIC)	CMR	Cameroon	Bamenda	141	Measured	132	NA	Measured
South-East Asia (LMIC)	IND	India	Agra	194	Measured	131	>75%	Measured

South-East Asia (LMIC)	IND	India	Gurgaon	124	Measured	120	50% -< 75%	Measured

Table 9: Top 10 Air polluted Cities In World 2019(WHO)

3.1.2.2 Urban emission :-

- According to Urban emission state of Bihar has been standing 3rd, 4th & 5th position in India from 1998 to 2020.
- The pie chart given below states that the major contribution in air pollution i.e (37.8%) is of All commercial and residential cooking, lighting, and heating & (17.5%) contribution is of industries.

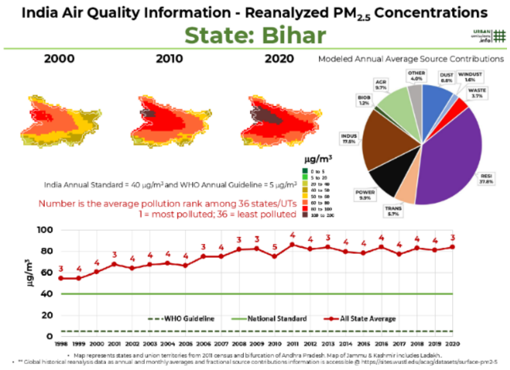


Figure 15:Bihar PM 2.5 Concentrations 2000-2020(Urban Emission)

3.2 Geographical features of Bihar

Bihar's geological composition is comprised of three major units: the Ganga Plain, Chotanagpur Plateau, and Himalayan foothills. The Ganga Plain is a flat alluvial plain in the northern and eastern regions of Bihar, formed by sediment deposition from the Ganges and its tributaries. The Chotanagpur Plateau, in the southern and western parts, is a hilly and mineral-rich area with coal, iron ore, bauxite, and copper deposits. The Himalayan foothills form the northern border, characterized by steep slopes, deep valleys, and natural hazards such as earthquakes and landslides. These diverse rock formations have played a significant role in shaping Bihar's natural resources and environmental conditions.

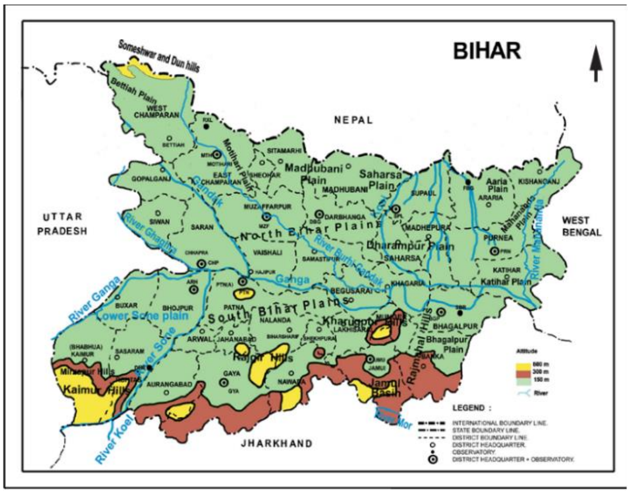


Figure 16 :PHYSICAL FEATURES OF BIHAR STATE ("[Climate of Bihar]" Climatological Summaries of States. No. 18. India Meteorological Department.)

3.2.1 Classification

According to the Köppen climate classification, Bihar's climate mainly falls under subtropical monsoon, mild and dry winter, and hot summer (Cwa), except for south-eastern parts of the state, such as Jamui , Banka, Munger, Lakhisarai, Khagaria, Shekhpura, some parts of Bhagalpur, Saharsa, and Begusarai. The southeastern part of the state is located in an extreme that falls under tropical savanna, hot, and seasonally dry (usually winter) (Aw)
Seasons

Winter

Bihar experiences cold weather from November to March, with pleasant temperatures in October and November. However, cold waves, known as Sheet-lahar locally, can cause sharp drops in temperature and disrupt the lives of millions of poor people, sometimes resulting in deaths. The temperature in winter varies from 0-10°C throughout Bihar, with December and January being the coldest months. In January 2013, several cities in Bihar recorded record lows, with Forbesganj reaching -2°C and Patna and Muzaffarpur reaching -1°C.

Summer

Hot weather arrives in March and lasts until the middle of June. The highest temperature is often registered in May. Like the rest of northern India, Bihar also experiences dust storms, thunderstorms, and dust-raising winds during the hot season. Dust storms with a velocity of 48–64 km/hour are most frequent in May, followed by April and June. The hot winds (loo) of the Bihar plains blow during April and May, with an average velocity of 8–16 km/hour. The hot wind greatly affects human comfort during this season.

Rain (monsoon)

Monsoon season in Bihar is usually unpredictable and erratic. It begins in mid-June and continues until the end of September.

Autumn (post-monsoon)

An important feature of the retreating monsoon season in Bihar is the invasion of tropical cyclones originating in the Bay of Bengal at about 12° N latitude. Bihar is also influenced by typhoons originating in the South China Sea. The maximum frequency of the tropical cyclones occurs in September– –November, especially during the asterism called hathiya. These cyclones are essential for the maturing of rice paddies and are required for the moistening of the soil for the cultivation of rabi crops.

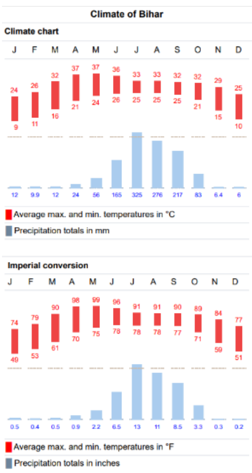


Figure 17 :Climate Chart of Bihar (Wikipedia & "Climate of Bihar" Climatological Summaries of States. No. 18. India Meteorological Department.)

3.3 Types Of Data Collection

1. Secondary Data
2. Primary Data

3.3.1 Secondary Data

3.3.1.1 Sources Of Secondary Data

Secondary data from the Central Pollution Control Board (CPCB) has been collected for the purpose of conducting a study on the air quality in the state of Bihar. This data will be utilized to analyze the current air quality status and identify any trends or patterns that may exist in the region, with the aim of providing valuable insights into the state of air quality in Bihar and contributing to the development of effective strategies for improving it.cpcb.nic.in

The first objective of our secondary data collection in this study is to identify the top five cities with the highest levels of air pollution in the state of Bihar. To achieve this, we utilized data obtained from the Central Pollution Control Board (CPCB) and calculated the Air Quality Index (AQI) for each station in Bihar from November 2016 to November 2022. After conducting a thorough analysis of the data, we identified the top five cities with the highest AQI readings. Our research will focus on further analysing the air quality in these five cities.So on the basis of Air quality index value we have got the most polluted substation in the State of Bihar.

- Data which shows most polluted substation

<table>
<tr><th>S.NO</th><th>STATION</th><th>SUBSTATION</th><th>AQI</th><th>SUBSTATION RANK</th></tr>
<tr><td>25</td><td>Siwan</td><td>Chitragupta Nagar, Siwan - BSPCB</td><td>293</td><td>1</td></tr>
<tr><td rowspan="3">18</td><td rowspan="3">Muzaffarpur</td><td>Muzaffarpur Collectorate</td><td>244.8571</td><td rowspan="3">2</td></tr>
<tr><td>Buddha Colony,</td><td>133</td></tr>
<tr><td>MIT-Daudpur Kothi, Muzaffarpur - BSPCB</td><td>275.5</td></tr>
<tr><td rowspan="6">19</td><td rowspan="6">Patna</td><td>IGSC Planetarium Complex</td><td>272.5714</td><td rowspan="6">3</td></tr>
<tr><td>Muradpur</td><td>170.25</td></tr>
<tr><td>Samanpura</td><td>180.25</td></tr>
<tr><td>Rajbansi Nagar</td><td>188</td></tr>
<tr><td>DRM Office Danapur</td><td>167.3333</td></tr>
<tr><td>Govt. High School Shikarpur, Patna - BSPCB</td><td>130.6667</td></tr>
<tr><td>11</td><td>Gaya</td><td>Collectorate, Gaya - BSPCB</td><td>201</td><td>4</td></tr>
</table>

		SFTI Kusdihra, Gaya - BSPCB	**57**	
		Kareemganj, Gaya - BSPCB	**267**	
8	**Buxar**	**Central Jail, Buxar - BSPCB**	**261.5**	**5**

Table 10: Top 5 Air polluted Cities In Bihar 2016-2022

Above give the order of most polluted substation (which represents stations) in Bihar state based on the AQI value from 2016 to 2022 data.
According to data Siwan is most polluted station followed by Muzaffarpur ,Patna ,**Gaya** , Buxar.
As we encountered insufficient data for the stations Buxar and Siwan, it was not feasible to conduct a time series analysis. Hence, our research work will be carried out on the remaining stations, namely Muzaffarpur, Patna, and Gaya, to proceed further with our study of air quality in the state of Bihar.

3.3.1.2 Secondary Research Design

Time series analysis is a specific way of analysing a sequence of observation collected over an interval of time. In time series analysis, one can record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly.

3.3.1.3 Classical method of forecasting

Component of time series

Decomposition methods assume that Yt, the actual time series value at period t, is a function of four components namely;
1. Trend component (T)
2. Seasonal component (S)
3. Cyclic component (C)
4. Irregular or error component (I)

Trend (T)

Trend is the long term movement of a time series. In other words, trend is the general tendency of time series to increase, decrease or stagnate.
Tendency of a time series is always not the same throughout the time period. There is a possibility of short-term up trends, down trends or even side line movements. But the overall long-term tendency may be upward, downward or stable. Trend can be linear or non linear. Trend is usually resulted by long term effects of factors such as socio-economic, political and demo-graphical characteristics.

Seasonal variation (S)

Short term variations in time series as a result of seasonal factors are defined as Seasonal Variation in a time series. Typically, seasonal variations occur within a specific period of time one year or shorter one year or shorter.
For an example, ice-cream sales of a shop may increase each year during summer, or sales of a toys store may increase every year during Christmas.

Time Series Data can also exhibit seasonal patterns of less than year in duration. For an example, daily traffic volume shows within the day “seasonal” behaviour, with peak levels occurring during rush hours.Weekly sales in a fashion store shows “weekly” seasonal behaviour where highest sales occur during weekend.

Cyclic variation (C)

Long term oscillations occurring in a time series are called cyclic variations. Term can be 5 years, 10 years or even more. These variations are mostly observed in economic and business data.
Most economic and business time series are influenced by wave like changes of prosperity and depreciation. There is periodic up and down movements and these movements occur in a cycle. Cyclic variation is extremely difficult to forecast. Hence, it is often combined with long-term trend effect.

Irregular variation (I)

Irregular Variations are fluctuations in time series that are short in duration, unpredictable in nature and follow no regularity in the occurrence pattern.
The irregular component is what is left after trend, seasonal variation and cyclic variation of a time series are estimated and removed. Hence, they are also known as ‘residual variations’.
Irregular variations can be either due to random effects or non-random effects such as earthquakes, tsunami or war.
Trend and cyclic variations are long term movements while seasonal and irregular variations are short term movements.

3.3.1.4 Methods of measuring Trend

Trend is measured using by the following methods:
a) Graphical method
b) Semi averages method
c) Moving averages method
d) Method of curve fitting

Method of curve fitting

Method of curve fitting is the method for finding the best fit of a set of data points. It minimizes the sum of the residuals of points from the plotted curves. It gives the trend line of best fit to a time series data. This method is most widely used in time series analysis.
Each point on the fitted curve represents the relationship between a known independent variable and an unknown dependent variable.
In general, the least squares method uses a straight line in order to fit through the given points which are known as the method of linear or ordinary least squares. This line is termed as the line of best fit from which the sum of squares of the distances from the points is minimized.
Equations with certain parameters usually represent the results in this method. The method of least squares actually defines the solution for the minimization of the sum of squares of deviations or the errors in the result of each equation.
The least squares method is used mostly for data fitting. The best fit result minimizes the sum of squared errors or residuals which are said to be the differences between the observed or experimental value and corresponding fitted value given in the model. There are two basic kinds of the least squares methods – ordinary or linear least squares and nonlinear least squares.

Merits

(i) The trend equation can be used to estimate or predict the values of the variable for any period t in future or even in the intermediate periods of the given series and the forecast values are also quite reliable.

(ii) The curve fitting by the principle of least square is the only technique which enables us to obtain the rate of growth per annum for year data if linear trend id fitted.

(iii) The higher order term tends to destroy the good fit toward the tail of the curve.

Demerits

(i) Future predication or forecast based on this method are best only on the long term variation i.e. trend and completely ignore the cyclical, seasonal and irregular fluctuation.

(ii) The most serious limitation of this method is the determination of the type of trend curve to be fitted i.e. whether we should fit a linear or a parabolic trend or some other more complicated trend curve.

(iii) It cannot be used to fit growth curve like modified exponential curve, Gompertertz curve and logistic curve and business time series conform.

3.3.1.4.1 Mathematical representation

It is a mathematical method and with it gives a fitted trend line for the set of data in such a manner that the following two conditions are satisfied.

- The sum of the deviations of the actual values of Y and the computed values of Y is zero.
- The sum of the squares of the deviations of the actual values and the computed values is least.

Using this method we compute parameters of lines which fits the data well. Followings are some of the commonly used models.

Straight line: $T_t = a+ bt$.

When the time series is found to be increasing or decreasing by equal absolute amount then we use this model.

Second degree polynomial: Tt=a0+a1t+a2t2.

We use this model when the data shows a curvature being concave either upward or downward.

Polynomial of order p: $T_t=a_0+a_1t+a_2t^2+\ldots+a_pt^p$.

Exponential curve: $T_t =ab^t$.

Just like second degree polynomial when the data shows a curvature being concave but the rate of change is way faster then we use this model.

Power curve: $T_t = at^b$

We use this model when the data shows a curvature being convex either upward or downward.

3.3.1.5 Forecasting by smoothing method

Smoothing is a technique applied to time series to remove the fine-grained variation between time steps. The hope of smoothing is to remove noise and better expose the signal of the underlying causal processes from which one can forecast. There are different methods of forecasting;

3.3.1.5.1 Holt's method of forecasting

This method is used when trend component present in the data it is just extension of exponential smoothing method for forecasting.
Let Mt denotes the mean level of the process at time t in the absence of trend component, and Tt denote the trend value at time t then the model will be

$$Yt = Mt\text{-}1 + Tt\text{-}1 + \varepsilon t$$

Where εt is the noise /error term with mean 0 and constant variance σ^2_ε
Then τ step ahead forecast of Yt is

$$Y_{t+\tau} = M_t + \tau T_t$$

3.3.1.6 Forecasting by Box Jenkins procedure

In time series analysis, the Box–Jenkins method, named after the statisticians George Box and Gwilym Jenkins, applies auto-regressive moving average or auto-regressive integrated moving average models to find the best fit of a time-series model to past values of a time series. It is a algorithm to find the fitted model and implement that.

Stationarity test

Stationarity means that the time series does not have a trend, has a constant variance, a constant auto correlation pattern, and no seasonal pattern. The auto-correlation function declines to near zero rapidly for a stationary time series. In contrast, the ACF drops slowly for a non-stationary time series.

Stationary of Time Series:

The time series {Xt, t≤∈Z} with index set Z= {0, ±1, ±2…...} is said to be stationary if
(i) $E(Xt) < \infty \ \forall \ t \in Z$
(ii) $E(Xt)=m \ \forall \ t \in Z$
(iii) $\gamma x(r,s)= \gamma x(r+t, s+t) \ \forall \ t \in Z$
Where γx(r,s) is auto-covariance function of the process {Xt, t≤∈Z}. This is also called weak Stationary or covariance stationary.

Test for stationary:

There are generally two methods to check stationary of a time series data.
(i) Graphical procedure (ii) ADF test

Graphical procedure:

This involves plotting the data over time and the corresponding (partial) auto-correlation function. If the ACF does not decrease to zero or at a very slow decay: this suggests non-stationarity (or long-memory effects).

Augmented Dickey-Fuller (ADF) test:

Statistical tests make strong assumptions about your data. They can only be used to inform the degree to which a null hypothesis can be rejected or fail to be rejected. The result must be interpreted for a given problem to be meaningful.
However, they provide a quick check and confirmatory evidence that the time series is stationary or non-stationary. The Augmented Dickey-Fuller test is also called a unit root test.
In probability theory and statistics, a unit root is a feature of some stochastic processes (such as random walks) that can cause problems in statistical inference involving time series models. In a simple term, the unit root is non-stationary but does not always have a trend component.
The hypothesis of ADF test is given as
Null Hypothesis (HO): Series is non-stationary or series has a unit root.
Alternate Hypothesis (HA): Series is stationary or series has no unit root.

If the null hypothesis is failed to be rejected, this test may provide evidence that the series is non-stationary.
If test statistic < Critical Value and p-value < 0.05 then, reject Null Hypothesis (HO) at 5% level of significance i.e., time series does not have a unit root, meaning it is stationary. It does not have a time-dependent structure.

Model identification
The ACF and PACF do have very distinct and indicative properties for MA and AR models, respectively. Therefore, in model identification, we strongly recommend the use of both the sample ACF and the sample PACF simultaneously.

Auto-correlation Function (ACF):-
Use the auto-correlation function (ACF) to identify which lags have significant correlations, understand the patterns and properties of the time series, and then use that information to model the time series data. From the ACF, you can assess the randomness and stationarity of a time series. You can also determine whether trends and seasonal patterns are present.

$$\rho k = (Xt,,Xt+k)/v(Xt)$$

Where v(Xt) is variance of Xt.

Partial Auto-correlation Function (PACF):-

The partial auto-correlation function is similar to the ACF except that it displays only the correlation between two observations that the shorter lags between those observations do not explain. For example, the partial auto-correlation for lag 3 is only the correlation that lags 1 and 2 do not explain. In other words, the partial correlation for each lag is the unique correlation between those two observations after considering partially out the intervening correlations.
As you saw, the auto-correlation function helps assess the properties of a time series. In contrast, the partial auto-correlation function (PACF) is more useful during the specification process for an auto-regressive model. Analysts use partial auto-correlation plots to specify regression models with time series data and Auto Regressive Integrated Moving Average (ARIMA) models. I'll focus on that aspect in posts about those methods.
Let {Xt, t≤∈Z} be a process then the partial auto correlation between Xt and Xt+h is defined as the conditional correlation between Xt and Xt+h condition on Xt-h ...Xt-(h-1) that is the set of observation in between time t and t-h the first order partial auto correlation is same as first order auto correlation the second order partial auto correlation is defined as

$$\frac{cov(x_t\,, x_{t-k}/x_{t-1}.., x_{t-(k-1)})}{\sqrt{var(x_t\,/x_{t-1}.., x_{t-(k-1)})}\sqrt{var(x_{t-k}/x_{t-1}.., x_{t-(k-1)})}}$$

Model identification
There are two basic time series models for the analysis of the stationary time series process. One is AR(p) and other is MA(q) model. The generalization of these two model is called ARMA(p,q) model which is the mixture of both models.
Auto regressive Processes of order p, AR(p)
The idea behind the auto-regressive models is to explain the present value of the series, Xt , by a function of p past values, Xt-1, Xt-2, . . ., Xt-p
An autoregressive process of order p is written as-

$$Xt = \varphi 1Xt\text{-}1 + \varphi 2Xt\text{-}2 + \ldots + \varphi pXt\text{-}p + Zt$$

where {Zt} is white noise, i.e., {Zt} ∼ WN(0, σ2), and Zt is uncorrelated with Xs foreach s < t.

Moving Average of order q, MA(q)
In time series analysis the moving-average model (MA model), also known as moving-average process, is a common approach for modelling uni-variate time series. The moving-average

model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.

The notation MA(q) refers to the moving average model of order q:

$$X_t=\mu+ Z_t+ \theta_1 Z_{t-1}+\ldots+ \theta_q Z_{t-q}= \mu+ \sum_{i=1}^{q} \theta_i Z_{t-1}$$

where μ is the mean of the series, the θ1, ..., θq are the parameters of the model and the εt, εt−1,..., εt−q are white noise error terms. The value of q is called the order of the MA model. This can be equivalently written in terms of the backshift operator B as

$$X_t=\mu+(1+\theta_1 B+\ldots+\theta_q B^q) Z_t$$

ARMA (p, q) Process

The process {Xt} is called ARMA (p, q) process if {Xt} is stationary and if for every t,

$$X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

Where {Zt} ∼ WN (0, σ2) and the polynomials (1 − φ1z − ... − φpzp) and (1+ θ1z + ... +θqZq) have no common zeros.

The process {Xt} is said to be an ARMA (p, q) process with mean μ if {Xt − μ} is an ARMA (p, q) process.

It is convenient to use the more concise form of

$$\varphi(B)X_t = \theta (B)Z_t,$$

where φ (·) and θ (·) are the pth and qth-degree polynomials

$$\varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p$$

&

$$\theta (z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

B is the backward shift operator (Bj Xt = Xt−j, Bj Zt = Zt−j, j = 0, ±1,...). The time series {Xt} is said to be an auto regressive process of order p (or AR (p)) if θ (z) ≡ 1, and a moving-average process of order q (or MA(q)) if φ(z) ≡ 1.

The ACF and PACF of ARMA(p, q) process are determined by AR(p) and MA(q) component respectively. It can therefore be shown that the ACF and PACF of ARMA(p, q) both shows exponential decay and/or damped sinusoid patterns, which makes the identification of the order of the ARMA(p, q) model relatively more difficult. A stationary time series model can be identify on the basis of the ACf and the PACF by using the given table as.

Model	ACF	PACF
AR(p)	Exponential decay and/or damped sinusoidal	Cuts off after larger p
MA(q)	Cuts off after lag q	Exponential decay and/or damped sinusoidal
ARMA(p, q)	Exponential decay and/or damped sinusoidal	Exponential decay and/or damped sinusoidal

Table 11:Model Identification Process

Parameter estimation

There are several methods such as the methods of moments, maximum likelihood, and least squares that can be employed to estimate the parameters in the tentatively identified model. However, unlike the regression models, most ARIMA models are nonlinear models and require the use of a nonlinear model fitting procedure.

Diagnostic testing

It is very important to test the residuals for any analysis. In time series we can use graphical and score based method. In graphical method, we simply plot the ACF and the PACF function of residual and check whether there is any significant spike on the plots. In score based method, we can use the following tests;

Box-Pierce test
The Box-Pierce Q-statistics are given by:

$$BP(k)=n \sum_{k=1}^{k} \rho_{a,k}^{2}$$

Where $\rho_{a,k}^{2}$ is the auto correlation coefficient at lag k.
n is the number of terms in different series;
K is the maximum lag being considered.

Ljung-Box test
The Lung-Box Q-statistics are given by :

$$LB(k)=n \times (n+2) \times \sum_{k=1}^{k} \times \frac{\rho_{a,k}^{2}}{n-k}$$

Where $\rho_{a,k}^{2}$ is the auto-correlation coefficient at lag k.
n is the number of terms in different series;
K is the maximum lag being considered.
The hypothesis of both the tests are defined as
Null Hypothesis (HO): The residuals are random.
Alternate Hypothesis (HA): The residuals are not random i.e. there is auto-correlation in residual for some lag.
If the null hypothesis is failed to be rejected, this test may provide evidence that the series is non-stationary.
If test statistic < Critical Value or p-value < 0.05 then, reject Null Hypothesis (HO) at 5% level of significance.

Model selection
It possible for the practitioner to consider several different models with various orders and compare them based on the model selection criteria such as AIC, AICC, and BIC. Another approach to model selection involves the use of information criteria such as AIC proposed by Akaike (1974a) or the Bayesian information criteria of Schwartz (1978). In the implementation of this approach, a range of potential ARMA models are estimated by maximum likelihood methods and for each model, a criterion such as AIC (normalized by sample size N, given by

$$AIC =- 2(\log - \text{likelihood}) + 2K$$

where, K is the number of model parameters.
or the related BIC given by

$$BIC =- 2(\log - \text{likelihood}) + d * \log(N)$$

where, N is the sample size of the training set, d is total number of parameters.
Lower the values of AIC and BIC better the model in terms of prediction.

Forecasting
Now, as we have selected the best consider model using residual diagnostic testing and model selection criteria and have estimated the parameters of the model we are well equipped to forecast.

3.3.2 Primary Data

3.3.2.1 Frame work to choose Gaya

3.3.2.1.1WHO

- The WHO data shows that among the top 15 polluted cities in the world 11 cities were from India which includes Kanpur, Faridabad , **Gaya**, Varanasi, Patna, Delhi , Lucknow ,Etc.[4]

Table 12:Top Air Polluted Cities In India 2019 (WHO)

Region	**iso3**	**Country**	**City/Town**	**PM10 Annual mean, ug/m3**	**note on converted PM10**	**PM2.5 Annual mean, ug/m3**	**Temporal coverage**	**note on converted PM2.5**
South-East Asia (LMIC)	IND	India	Kanpur	(319)-converted value	Converted	173	>75%	Measured
South-East Asia (LMIC)	IND	India	Faridabad	(316)-converted value	Converted	172	>75%	Measured
South-East Asia (LMIC)	IND	India	**Gaya**	(275)-converted value	Converted	149	50% -< 75%	Measured
South-East Asia (LMIC)	IND	India	Varanasi	260	Measured	146	>75%	Measured
South-East Asia (LMIC)	IND	India	Patna	(266)-converted value	Converted	144	>75%	Measured

3.3.2.1.2 By Secondary Data

Our primary data analysis will focus on Gaya, as it was among the top five polluted cities in Bihar based on the air quality index calculated from the secondary data obtained from the Central Pollution Control Board. According to WHO data, 11 out of the top 15 polluted cities in the world are located in India, including Gaya, which makes it imperative to assess the air quality and its impact on public health. Furthermore, as students of the Central University of South Bihar, we have a particular interest in understanding the status of air quality in Gaya and its effect on the health of its residents, given that it is a city in Bihar where we are pursuing our Masters degree. Thus, our research methodology will involve collecting primary data from Gaya to analyze the air quality and assess its impact on public health.

3.3.2.2 Sample size Determination

Determination Of Sample Size:- The sample size determination is done in two steps viz. 1]Calculating the sample size for an infinite population.

2] Adjusting the sample size for the finite population.

So we proceed as:-

1]CALCULATING SAMPLE SIZE FOR INFINITE POPULATION:-

The Cochran's Formula for the determination of sample size for an infinite population as below

S = [((Z^2)* p * (1-p)) / (e)^2]

where,

S = Sample size for infinite population
Z = Z-score for the corresponding confidence interval
p = Population proportion= 50% effect of air pollution on people staying in Gaya.
e = Margin of error fixed by researcher
So the value of S for the following values :- Z = 1.96, p = 0.5, e = 0.05, comes out to be
S = 384
2]ADJUSTING SAMPLE SIZE FOR FINITE POPULATION:-
We use the following formula for adjusting sample size for a finite population:-
S1 = S / (1 + ((S - 1)/(P0)))
where;
S1 = Adjusted sample size for finite population
S = Sample size for Infinite population
P0 = Finite population (total population of Gaya according to the census 2011)
So the value of S1 for the following values :- S = 384, P0 = 468,614 comes out to be S1 = 384
Hence, we have collected our samples by using further given data collection method

3.3.2.3 Method Of Data Collection

Since our sampling unit is residents, working people, and students in Gaya. We have collected primary data using the Questionnaire method comprising questions on personal information, health effects, awareness & opinion about Air pollution in Gaya.

3.3.2.3.1 Questionnaire Method:

A questionnaire refers to a technique of data collection that consists of a series of written questions along with alternative answers. We choose to adopt the questionnaire method as the most effective method of data collection.

3.3.2.4 Sampling Technique :

3.3.2.4.1 Stratified Random Sampling:

- The procedure of partitioning a given population into homogeneous groups, called strata, and then selecting samples independently from each stratum is known as stratified sampling. If a sample from each stratum is selected by random sampling, the procedure will be called stratifiedrandom sampling
- In the first stage, we have divided Gaya into 3 zones which consist of three stations monitoring air quality representing the Gaya district according to CPCB. Here, we divided wards of Gaya as per the information given by the census 2011 into 3 zones on a directional basis as East, West & South zones. Each zone was considered a stratum. Administrative convenience was considered as a basis for stratification. Further, by using a stratified random sampling technique we randomly selected samples from each stratum.
- The zones are divided accordingly :

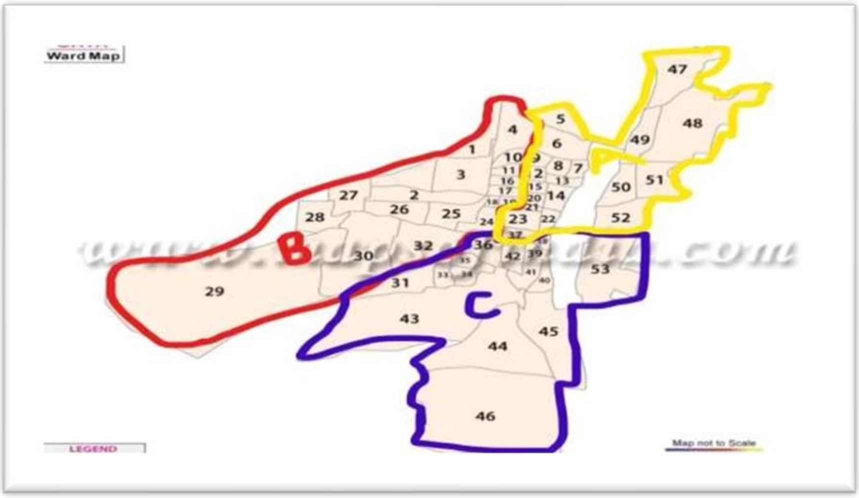


Figure 18 :Allocation Of Zones On Directional Basis

The zones are divided accordingly :

	ZONE A		ZONE B		ZONE C		
SR.NO.	WARD NO.	POPULATION	WARD NO.	POPULATION	WARD NO.	POPULATION	
1	5	3734	1	9398	31	8804	
2	6	13776	2	8264	33	5195	
3	7	8676	3	12881	34	10865	
4	8	10714	4	13015	35	4990	
5	9	5741	10	8219	36	12581	
6	12	10526	11	6779	38	10574	
7	13	4134	16	6901	39	5535	
8	14	10474	17	6422	40	8584	
9	15	7279	18	4365	41	6747	
10	20	4576	19	5139	42	7399	
11	21	2700	24	6636	43	10175	
12	22	8869	25	10346	44	5800	
13	23	7187	26	7003	45	12501	
14	37	3678	27	15984	46	9344	
15	47	7974	28	10663	53	11628	
16	48	13521	29	8407			
17	49	15482	30	15981			
18	50	15587	32	7402			
19	51	8487					
20	52	10972					**TOTAL**
TOTAL	**174087**		**163805**		**130722**		**468614**

Table 13:Sample Size Determination According To Zone Wise

Further, by using the proportional allocation method we determined the sample size for each zoneas:

Sample Size for Each Zone By Using Proportional Allocation			
ZONES	**POPULATION**	**SAMPLE SIZE**	**APPROX.**
ZONE A(n1)	**174087**	**142.6534589**	**143**
ZONE B(n2)	**163805**	**134.228**	**134**
ZONE C(n3)	**130722**	**107.1185411**	**107**
TOTAL	**468614**	**384**	**384**

Table 14:Sample Size For Each Zone By Using Proportional Allocation

Hence, we have collected a total of 367 samples from 3 zones

3.3.2.5 Data Analysis Techniques To prove Primary Objectives

3.3.2.5.1 Graphical Representation Techniques

1. Pie Chart:

Definition: A pie chart is a circular chart that shows the proportion of each category in a data set. Each category is represented by a slice of the pie.
Usage: Pie charts are commonly used to show the percentage breakdown of a data set, particularly when there are only a few categories.
Example: A pie chart could be used to show the percentage of different types of fruits sold in a grocery store, with each type of fruit represented by a slice of the pie.

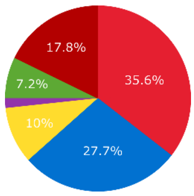


Figure 19 :Pie Chart

2. Donut Chart:

Definition: A donut chart is a circular chart similar to a pie chart, but with a hole in the center. Each category is represented by a slice of the donut.
Usage: Donut charts are commonly used to show the same information as pie charts, but with the added benefit of being able to show multiple data sets on the same chart.
Example: A donut chart could be used to show the percentage breakdown of different types of expenses in a household budget, with each type of expense represented by a slice of the donut.

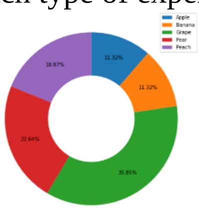


Figure 20 :Donut Chart

3. Bar Chart:

Definition: A bar chart is a chart that shows the comparison between different categories or groups. The x-axis represents the categories or groups, and the y-axis represents the values or counts.
Usage: Bar charts are commonly used to show the differences or similarities between different groups or categories.
Example: A bar chart could be used to show the average grades of different classes in a school, with each class represented by a bar.

Figure 21 :Bar Chart

4. Multiple Bar Chart:

Definition: A multiple bar chart is a chart that shows the comparison between multiple categories or groups. The x-axis represents the categories or groups, and the y-axis represents the values or counts.
Usage: Multiple bar charts are commonly used to show the differences or similarities between multiple groups or categories.
Example: A multiple bar chart could be used to show the average grades of different classes in a school, with each class represented by a set of bars for each subject.

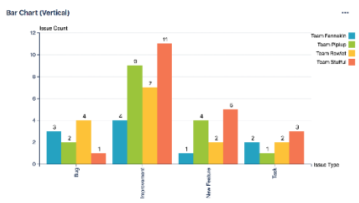


Figure 22 :Multiple Bar Chart

3.3.2.5.2 Statistical Tests

1. Chi-Square Test:

The Chi-square test is a statistical method used to test for independence between two categorical variables. The test assumes that the variables being tested are categorical and have a limited number of mutually exclusive categories. The null hypothesis for the test is that there is no association between the variables, while the alternative hypothesis is that there is a significant association.
The formula for the Chi-square test statistic is:
$\chi^2 = \Sigma(O - E)^2 / E$
where:
χ^2 is the Chi-square test statistic
O is the observed frequency

E is the expected frequency

Assuming a large enough sample size, the test statistic follows a Chi-square distribution with (r-1)(c-1) degrees of freedom, where r is the number of rows and c is the number of columns in the contingency table.

2. Cross-Tabulation:

Cross-tabulation is a method of summarizing and analyzing the relationships between two or more categorical variables. The method involves creating a contingency table that displays the frequency of observations for each combination of the categorical variables.
Assuming that the data are categorical, the method makes no specific assumptions about the underlying distribution of the data. However, it does assume that the categories being used are mutually exclusive and exhaustive. Cross-tabulation can be used to explore relationships between variables, identify subgroups within the data, and test hypotheses about the relationships between variables.
Certainly, here is an explanation of binary logistic regression in the format you requested:

3. Binary logistic regression

Definition:
Binary logistic regression is a statistical method used to model the relationship between a binary dependent variable and one or more independent variables. The dependent variable takes on two possible values, typically represented as 1 and 0, indicating the presence or absence of an outcome or characteristic of interest. The independent variables can be continuous, categorical, or a combination of both.
Formula:

The logistic regression model is based on the logistic function, also known as the sigmoid function, which maps the linear combination of the independent variables to a probability value between 0 and 1. The logistic function is defined as follows:

$$P(Y=1|X) = 1 / (1 + e^{-(B0 + B1X1 + B2X2 + ... + BkXk)})$$

where:

- → P(Y=1|X) is the probability of the dependent variable Y taking on the value 1, given the values of the independent variables X.
- → B0, B1, B2, ..., Bk are the coefficients of the independent variables X1, X2, ..., Xk, respectively.
- → e is the natural logarithm base.

Assumptions:

The logistic regression model assumes that:
- → The dependent variable is binary and follows a Bernoulli distribution.
- → The relationship between the independent variables and the dependent variable is linear on the logit scale.
- → The independent variables are independent of each other.
- → There is no multicollinearity among the independent variables.
- → The observations are independent of each other.

Uses:

Binary logistic regression is commonly used in various fields, such as:
- → Medicine: to study the risk factors for diseases and the effectiveness of treatments.
- → Psychology: to study the determinants of behaviors and the prevalence of disorders.

→ Sociology: to study the factors that influence social phenomena, such as voting behavior and crime rates.

Overall, binary logistic regression is a valuable statistical method that allows researchers to examine the relationship between a binary dependent variable and one or more independent variables, and to estimate the probability of the dependent variable taking a particular value.

4. Multinomial regression

Multinomial regression is a statistical method used to analyze categorical dependent variables with more than two levels. It is an extension of binary logistic regression and is commonly used in social and medical research.

The formula for the multinomial regression model is:

$$\ln(p_{ij} / p_{im}) = \beta_{0j} + \beta_{1j}X_{1i} + \beta_{2j}X_{2i} + ... + \beta_{pj}X_{pi}$$

Where p_{ij} is the probability of observation i belonging to category j, p_{im} is the probability of observation i belonging to the reference category m, β_{0j} is the intercept for category j, β_{kj} is the coefficient for independent variable X_k in category j, and X_{ki} is the value of independent variable k for observation i.

Assumptions of multinomial regression include:

1. The dependent variable is categorical with more than two levels.
2. The observations are independent.
3. There is no multicollinearity among the independent variables.
4. The error terms are independent and have a constant variance.
5. The relationship between the independent variables and the dependent variable is linear.

Multinomial regression is used in a variety of fields, such as marketing, political science, and healthcare. It is used to model and predict the probability of an outcome falling into one of several categories. For example, in healthcare, it could be used to predict the likelihood of a patient having one of several diseases based on their symptoms.

In summary, multinomial regression is a powerful statistical tool used to analyze categorical data with more than two levels. Its formula, assumptions, and uses make it a popular choice for researchers in many fields.

5. Odds Ratio Definition:

The odds ratio is a measure of the strength of the association between an exposure and an outcome in epidemiological research. It compares the odds of the outcome occurring in the exposed group to the odds of the outcome occurring in the unexposed group.

Formula:

The odds ratio (OR) is calculated as:

OR = (a/b) / (c/d)

where:

→ a is the number of exposed individuals who develop the outcome of interest
→ b is the number of exposed individuals who do not develop the outcome of interest
→ c is the number of unexposed individuals who develop the outcome of interest
→ d is the number of unexposed individuals who do not develop the outcome of interest

Interpretation:

If the odds ratio is greater than 1, it suggests that the exposure is associated with an increased risk of the outcome. If the odds ratio is less than 1, it suggests that the exposure is associated with a decreased risk of the outcome. If the odds ratio is equal to 1, it suggests that the exposure is not associated with the outcome.

CHAPTER 4 : FORECASTING PM2.5 CONCENTRATION USING TIME SERIES MODELING IN A POLLUTED REGION OF BIHAR

4.1 Background

The air pollution data for the period 2016-2022 is available at the central pollution control board site. Here we have used the data from the period 2016-2022 in our study which contain daily air quality data of 36 air quality monitoring stations in State Of Bihar.

4..2 Data link

Here we have considered the secondary data for our analysis. The source of the original data is given from the following website.Cpcb.nic.in

4.3 Data processing

Daily data is available in the form of (.xlsx) i.e. Excel format.
We have converted the given available daily data into monthly average & then by taking average of monthly data we converted it to yearly scale in excel. Following is the data of monthly & yearly average of parameter PM2.5 of 3 stations that we have selected on the basis of worst the air quality index.

Status of Ambient Air Quality of Gaya (Collectorate)								
Main pollutants								
SR.NO.	**Month**	**PM2.5 in µg/m3**						
		2016	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
1	**January**	164.466129	265.6384	102.2148148	197.9464516	96.41677419	82.30290323	77.35035714
2	**February**	88.57448276	318.75542	103.9212005	126.6805673	82.68241379	87.68458287	66.00932266
3	**March**	58.584	192.5014286	140.0448387	79.15666667	51.38966667	58.89222222	60.53483871
4	**April**	61.68137931	174.7643333	148.4233333	61.39275862	45.91192473	68.04233333	59.13366667
5	**May**	134.3270968	97.65612903	84.17	61.79774194	43.5364	49.67709677	47.00814815
6	**June**	111.8808333	104.0533333	69.5352	62.706	61.18	30.78466667	45.6804
7	**July**	39.993	124.85	35.589	44.54903226	61.18	40.20677419	41.82806452
8	**August**	28.05032258	205.2909211	34.92354839	40.61366667	28.33875	31.25903226	41.04172414
9	**September**	106.4275	205.2909211	46.20833333	31.62068966	44.30619048	31.57233333	36.31833333
10	**October**	247.4647059	276.94	75.92806452	62.46516129	61.18	55.80032258	60.62548387
11	**November**	243.8003333	184.3943333	139.0996667	109.2583333	67.04346154	96.12884615	108.3683333
12	**December**	348.8532258	313.3558333	169.9706452	121.3380645	90.99709677	101.8413043	161.5
	Avg	**136.1752507**	**205.2909211**	**95.83572045**	**83.29376115**	**61.18022318**	**61.1827015**	**67.11655604**

Table 15 : Status of Ambient Air Quality of Gaya (Collectorate)

Status of Ambient Air Quality of Patna (IGSC Planetarium)								
Main pollutants								
SR.NO.		PM2.5in µg/m3						
	Month	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
1	**January**	353.3629032	207.088	212.1864516	263.7009677	188.3522581	194.0832258	167.4854839
2	**February**	200.8792	175.9458333	148.0025	140.7128571	118.6676	167.3439286	128.08
3	**March**	130.4846667	121.0662069	119.2922581	86.64580645	76.95444444	86.13125	111.69
4	**April**	78.99433333	116.695	89.95466667	53.475	67.716	80.19758621	94.35966667
5	**May**	87.43387097	71	70.63225806	56.3916129	100.961107	47.25269231	56.76870968
6	**June**	63.79896552	60.92307692	66.674	45.16	41.14090909	45.758	51.043
7	**July**	43.71064516	49.11111111	44.2116129	31.97290323	38.51041667	37.54392857	32.68225806
8	**August**	51.15677419	40	53.42774194	33.99774194	35.05677419	28.58727273	31.92258065
9	**September**	61.884	47.6525	63.20533333	39.268	50.06966667	105.5418243	39.40133333
10	**October**	144.0903448	43.31206897	108.0883871	101.173	93.50290323	105.5418243	90.86
11	**November**	252.596	195.6306897	211.386	201.6946667	161.0773333	198.6377778	146.335
12	**December**	258.4190323	252.0270968	259.1009677	217.1154839	239.523871	169.8825806	224.8064516
	Avg	**143.9008947**	**115.037632**	**120.5135148**	**105.9423367**	**100.961107**	**105.5418243**	**97.95287366**

Table 16 : Status of Ambient Air Quality of Patna (IGSC Planetarium)

Status of Ambient Air Quality of Muzaffarpur (Collectorate)								
Main pollutants								
		PM2.5 in µg/m3						
	Month	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
1	**January**	290.717931	221.839	220.374	223.1926667	200.9474194	207.8874194	174.8725
2	**February**	179.4634483	224.8453571	166.7835714	129.9875	131.6517241	177.6328571	115.9810526
3	**March**	106.0176923	99.97967742	99.50548387	98.69709677	79.04741935	120.9458065	114.9612903
4	**April**	101.2517857	87.75962963	63.40793103	56.87933333	46.178	89.4137931	83.08733333
5	**May**	49.55107143	92.40074074	44.35129032	59.63096774	30.51	42.32354839	42.43483871
6	**June**	25.92103448	44.88190476	42.41433333	34.06666667	21.07833333	38.42733333	38.83689655
7	**July**	19.62827586	26.625	22.3273913	102.12766	18.15193548	27.57421053	25.67032258
8	**August**	32.15741935	27	24.1576	33.63777778	14.93516129	23.13083333	26.00677419
9	**September**	32.83966667	112.885	41.92282581	24.78033333	23.865	29.58551724	29.35433333
10	**October**	129.4319355	13.53565217	93.44451613	87.47774194	70.26032258	64.26107143	53.51333333
11	**November**	208.898	176.6731884	217.1523333	182.9793333	167.9386667	164.2958621	115.4165517
12	**December**	247.4480645	248.4703226	244.2813793	192.0748387	250.2132258	168.7048387	164.5319231
	Avg	**118.6105271**	**114.7412894**	**106.676888**	**102.1276597**	**87.89810067**	**96.18192426**	**82.05559582**

Table 17 : Status of Ambient Air Quality of Muzaffarpur (Collectorate)

4.4 Study area

As per the available data for the year 2016-2022 provided by the central pollution control board & after finding the most polluted cities in the State Of Bihar we got three stations which are as follows:-

Serial No.	Station name	Data period	Time scale
1	Gaya	2016-2022	Daily data
2	Patna	2016-2022	Daily data
3	Muzaffarpur	2016-2022	Daily data

Table 18 : Station Wise Data Availability Period

The location of these three stations in map is given below:-

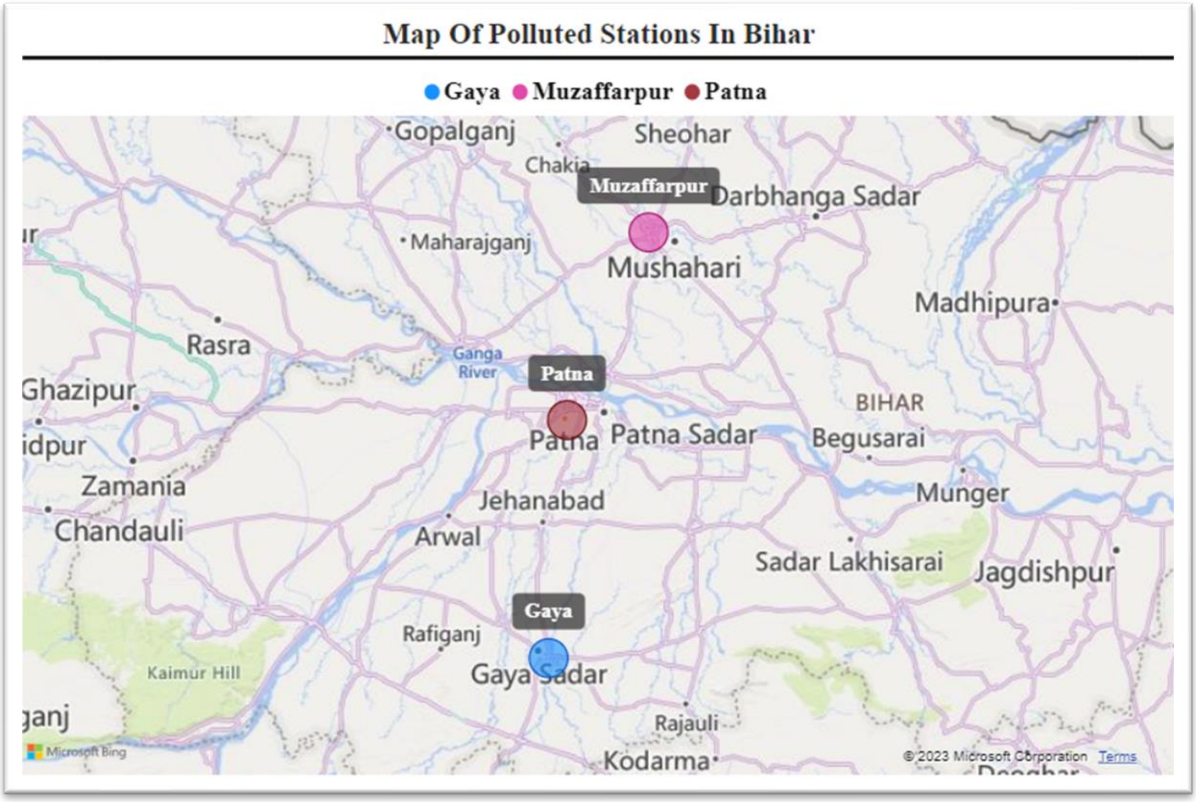


Figure 23 : Map of selected polluted stations

4.5 Study objectives

- The 5 most polluted cities in Bihar
- Forecasting Pm2.5 concentration for the year 2022 & 2023 in a polluted region of Bihar by using
  - → Trend method
  - → Smoothing method
  - → Model fitting

4.6 Analysis

The analysis was conducted on the yearly average data of the PM2.5 parameter from 2016 to 2021 for three stations in Bihar, using the data provided by the Central Pollution Control Board. PM2.5 was considered as the major parameter that influences air quality.

Initially, we created separate graphs for each of the three stations to plot the yearly average data of the PM2.5 parameter from 2016 to 2021.

Station 1: Gaya

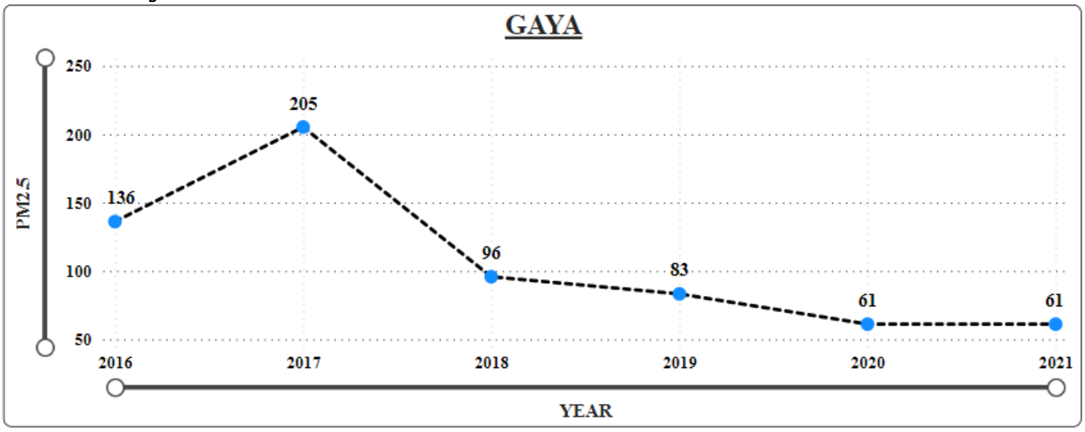


Figure 24 : Yearly average value of parameter PM2.5 in Gaya from 2016-2022

Station 2: Patna

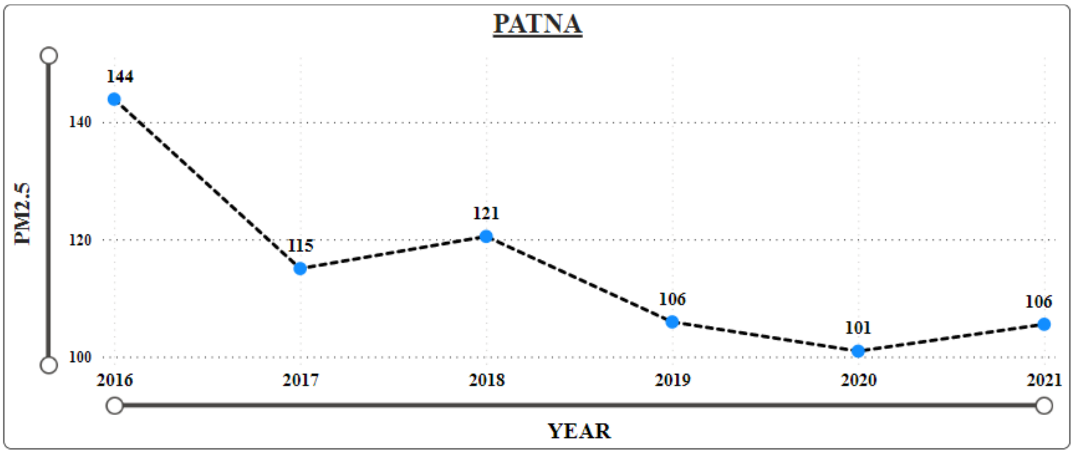


Figure 25 : Yearly average value of parameter PM2.5 in Patna from 2016-2022

Station 3: Muzaffarpur

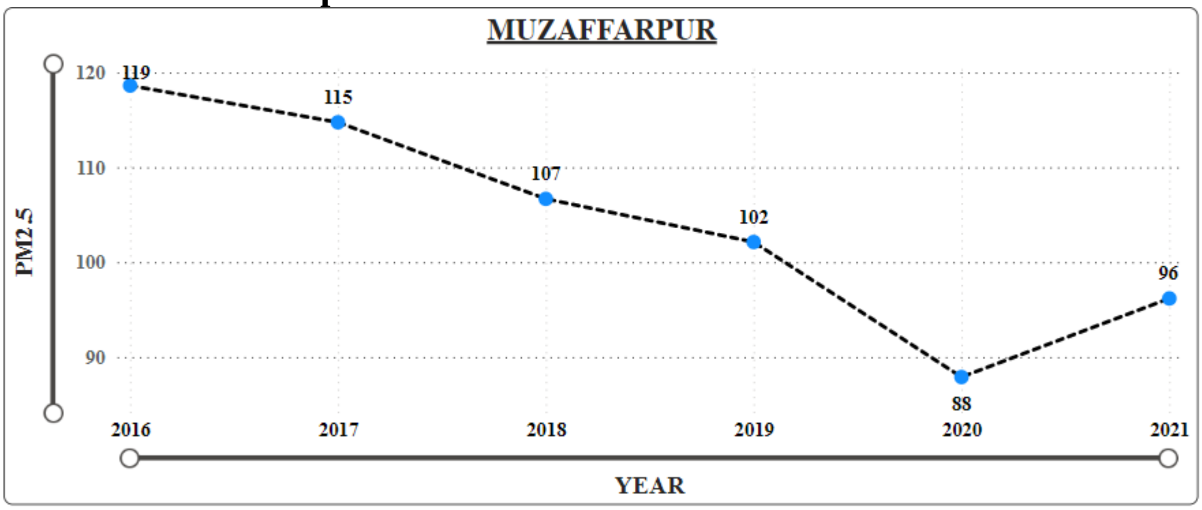


Figure 26 : Yearly average value of parameter PM2.5 in Muzaffarpur from 2016-2022

- The urban emission data shows that Bihar has been consistently ranked 3rd, 4th, and 5th in India from 1998 to 2020.
- As per the data provided by the Central Pollution Control Board (CPCB), the cities of Gaya, Patna, and Muzaffarpur have the worst air quality in the state of Bihar.
- Each station's values have shown a decreasing trend year-on-year, indicating a possibility of improvement in air quality for the three stations in the coming years. If we can forecast the values for the years 2022 and 2023, we may be able to draw some conclusions about the air quality.
- Upon analysing the PM2.5 parameter graphs for the reported years from 2016 to 2021, we proceeded to forecast the values for 2022 and 2023.
- The curve fitting method was employed using the Power, logarithmic, and linear methods to forecast the trend value.
- The smoothing method was employed, with Holt's method utilized for smoothing. Additionally, we attempted to forecast the average PM2.5 parameter values for 2022 and 2023 through a time series model, utilizing Box-Jenkins' procedure.
- In Box-Jenkins' procedure, we first checked for the stationarity of the data. We employed the ADF test to test the stationary nature of the considered dataset. If the data does not display stationary behaviour by the test, we transform it by considering the successive differences at different orders to make it stationary. In such cases, we use the ARIMA (p,d,q) process, where the order specifies the value of d in the ARIMA (p,d,q) process.
- After confirming data stationarity, the next step involves process identification by utilizing ACF and PACF plots. During identification, all possible models and values of (p,q) in the ARMA (p,q) process are considered. The parameters involved in all selected models are then estimated using any available estimation methods. Next, the residuals are analysed to determine whether they are random or not for each considered model, using methods such as Box-Pierce's test and Ljung-Box test. Any model failing these tests is discarded, and the remaining models undergo model selection using Bayesian information criterion (BIC). The model with the minimum BIC is deemed most suitable and utilized for forecasting the average PM2.5 parameter value.

4.6.1 Analysis of data for station 1 (Gaya)

4.6.1.1 Trend (Curve Fitting Method)

In order to analyse & forecast the average value of parameter for the station Gaya we have used power trend & fitted trend line which is shown in the given figure.

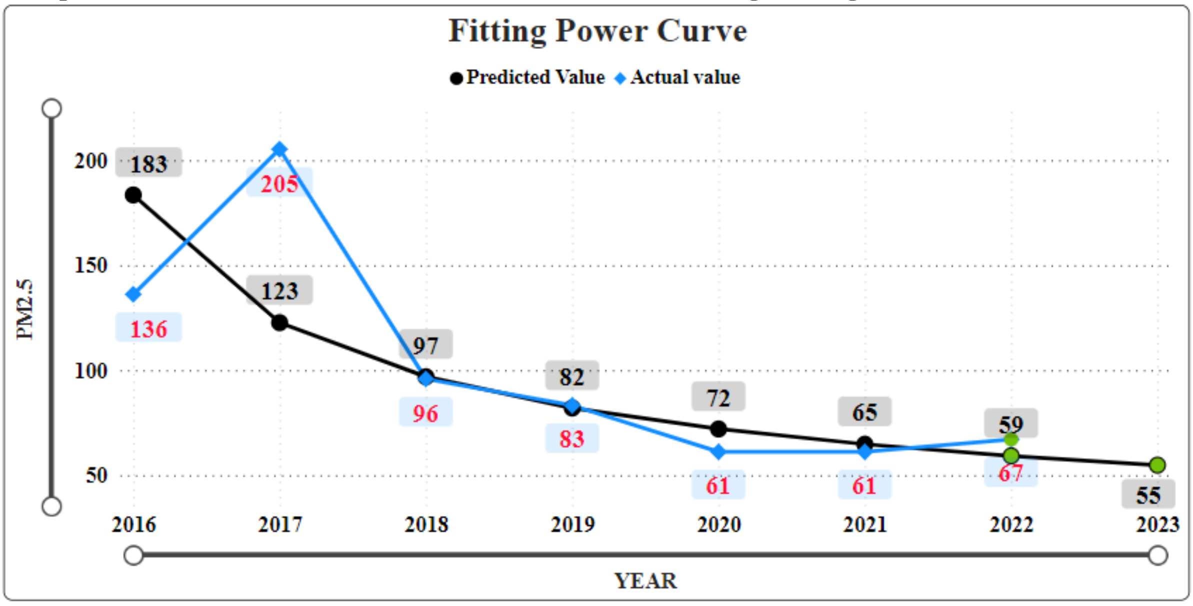


Figure 27 : Fitting Power trend curve on parameter PM2.5 for the station Gaya

The above graph shows a power trend curve fitted on the average value of the parameter PM2.5 for the years 2016-2021, which was collected from the central pollution board site. This curve was used to forecast the value for the year 2022, which was then compared to the actual value of 2022. Furthermore, a predicted value for the year 2023 has also been generated using this power trend curve.

Year	Actual	Predicted
2022	67	59
2023		55

Table 19 : Prediction For Year 2022-23 By Power Trend (Gaya)

The equation we get for this power curve trend is
$y = 183.49x^{-0.581}$; where x is year

4.6.1.2 Forecast using Holt's method

Based on the trend observed in our data, we have utilized Holt's method for forecasting. For this purpose, we have chosen the smoothing constants α and β as 0.8 and 0.25, respectively. By applying these values to the data collected from 2016 to 2021, we have made a forecast for the year 2022, which we then compared to the actual value for 2022. Additionally, we have used this method to forecast the value for the year 2023.

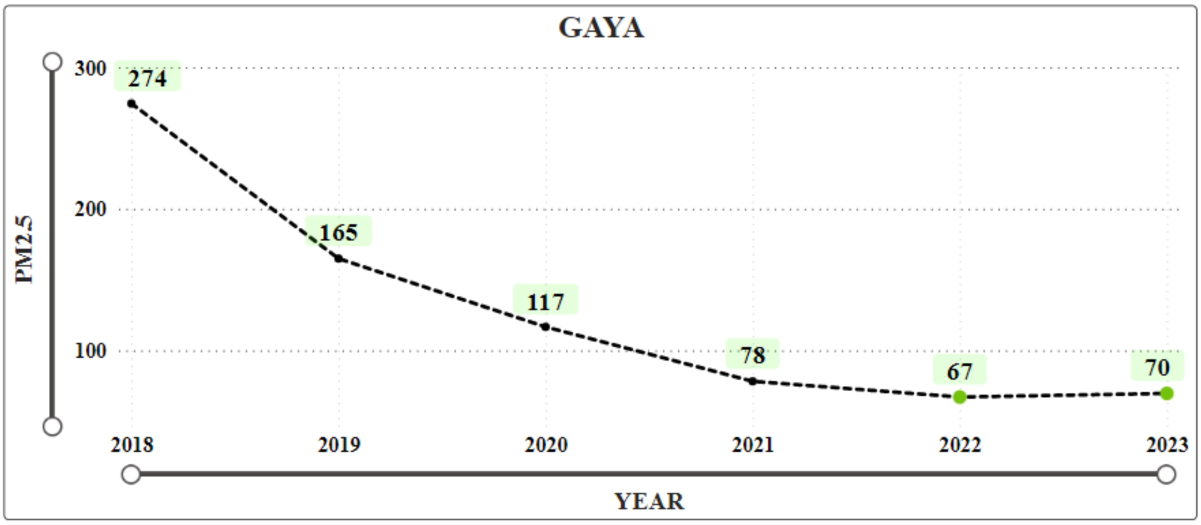


Figure 28 : Forecasting by Holt's method for Gaya station

The predicted value we got from Holt's method is

Year	Actual	Predicted
2022	67	67.1496
2023		69.7031

Table 20 : Prediction For Year 2022-23 By Holt's Method (Gaya)

4.6.1.3 Forecast using Box-Jenkins procedure

In the past, we used classical methods and smoothing techniques to forecast the value of the parameter PM2.5 for the years 2022 and 2023 using data from 2016 to 2021. However, we have since utilized the Box-Jenkins procedure to determine the best-fitted model for our data from 2016 to 2022.

Using this model, we have forecasted the value of PM2.5 for the year 2023. Therefore, our prediction for the PM2.5 value in 2023 is based on the Box-Jenkins model applied to the data from 2016 to 2022. This approach offers a more sophisticated and accurate means of forecasting the value of the parameter, as it takes into account the complex relationships and patterns present in the data.

Test for stationarity

We have used augmented Dickey-Fuller test to check the stationarity.
Here the null hypothesis is the data is non-stationary and the alternative hypothesis is the data is stationary

Augmented Dickey-Fuller Test

Null: non-stationary
Alternative: stationary

Lag	ADF	P-value
1	-0.859	0.43

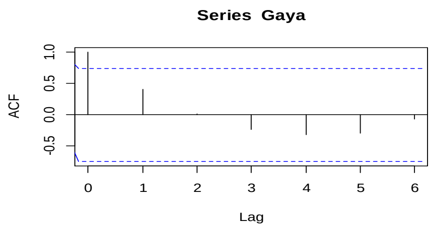
Table 21 : Augmented Dickey-Fuller Value & P-Value (Gaya)

Initially, we performed the ADF (Augmented Dickey-Fuller) test on our data and found that the p-value for lag 1 was greater than 0.05, which means that we could not reject the null hypothesis at 5% level of significance based on our data. This indicated that our data was non-stationary.

To address this issue, we took the first-order difference of the data and performed the ADF test again. The result of the first difference data showed that the p-value for the lag 1 was now less than 0.05, indicating that we could reject the null hypothesis at 5% level of significance. This suggested that the first-order differenced data was stationary. Therefore, we can conclude that taking the first-order difference of the data made it stationary and suitable for further analysis & the model for our data will be ARIMA with d=1.

Model identification

As we have already discussed that, for model identification, we use ACF and PACF plots. These figures shows the ACF and PACF plot for station 1



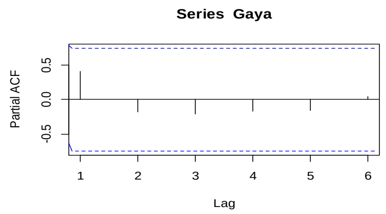


Figure 29 : ACF & PACF of air pollutant parameter PM 2.5 for the Gaya station

We can see that the above ACF plot shows a damped sinusoidal pattern but has no cut points.
Here we can see that the above PACF plot has no cut points so, the suitable model may as ARIMA(0,1,1) Or ARIMA(1,1,0) model.
After considering suitable models for the station 1, we followed the procedure of Box-Jenkins' for each of the model one by one

Parameter estimation of ARIMA(1,1,0) & ARIMA (0,1,1) model

MODELS	LOG L	φ1	θ1	θ2	σ2	K	AIC	BIC
ARIMA(0,1,1)	-32.1		-0.3081		2551	2	68.19	68.09182
ARIMA(1,1,0)	**-32.02**	**-0.3815**			**2461**	**2**	**68.04**	**67.93182**

Table 22 : Parameter Estimation

BIC of ARIMA(p,d,q) model

We use Bayesian Information Criterion to select one of two models. Here for ARIMA(1,1,0) we got a BIC score of 67.9318 which is less than the BIC score i.e.68.0918 for ARIMA(0,1,1) model.

Box-Pierce test

The result of Box-Pierce test for the model ARIMA(1,1,0) is shown below
X-squared = 0.07302, df = 1, p-value = 0.787
From the score based approach we can see that p value is greater than 0.05 which means that the residuals are randomly distributed which means we can use this model for forecasting.

Forecasting using ARIMA(1,1,0) model

Here we have forecasting air pollutant parameter PM2.5 value for the year 2023 using data from 2016-2022 with the model ARIMA(1,1,0)

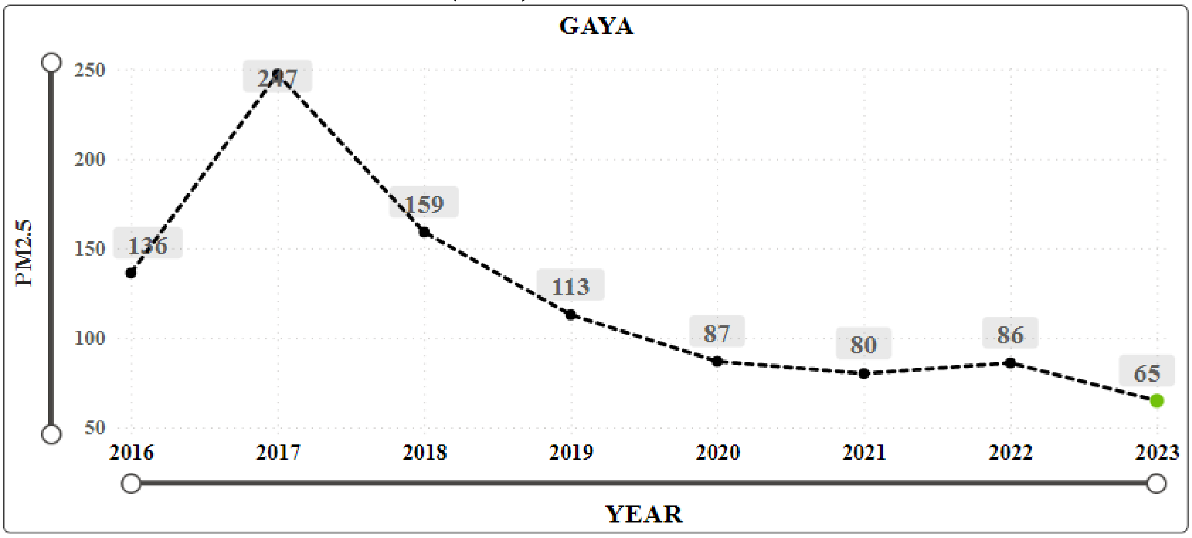


Figure 30 : Forecasting using ARIMA(1,1,0) model for Gaya station

Following is the forecast value

Year	Predicted Value	Standard Error
2023	64.85306	49.6045

Table 23 : Forecast Value

4.6.2 Analysis of data for station 2 (Patna)

4.6.2.1 Trend (Curve Fitting Method)

In order to analyse & forecast the average value of parameter for the station Patna we have used Logarithmic trend & fitted trend line which is shown in the given figure.

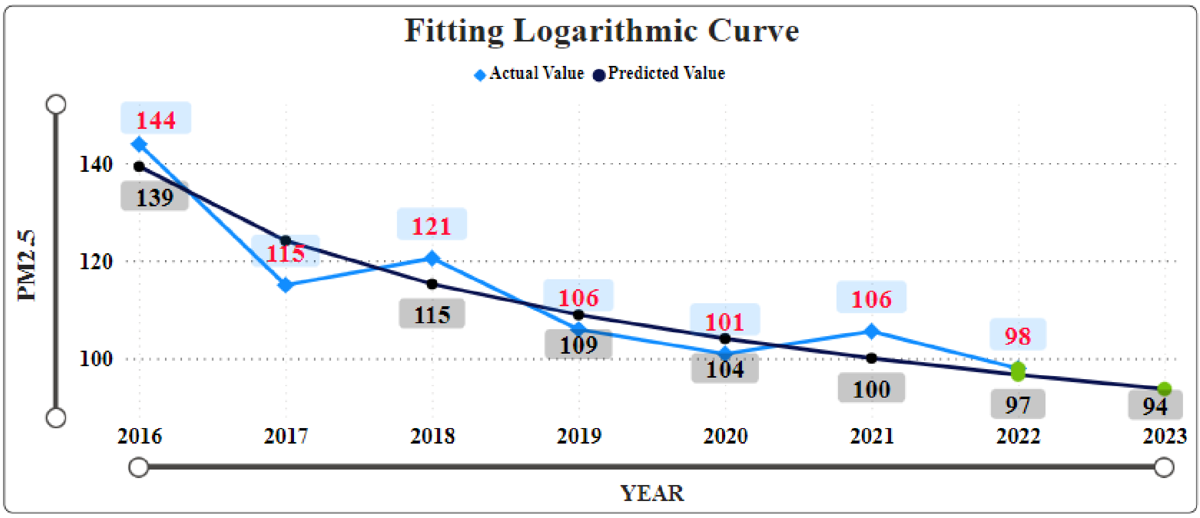


Figure 31 : Fitting Logarithmic curve on parameter PM2.5 for the station Patna

The above graph shows a logarithmic trend curve fitted on the average value of the parameter PM2.5 for the years 2016-2021, which was collected from the central pollution board site. This curve was used to forecast the value for the year 2022, which was then compared to the actual value of 2022. Furthermore, a predicted value for the year 2023 has also been generated using this logarithmic trend curve.

Year	Actual	Predicted
2022	98	97
2023		94

Table 24 : Prediction For Year 2022-23 By Power Trend (Patna)

The equation we get for this logarithmic curve is
$y = -21.91*\ln(x) + 139.34$; where x is year

4.6.2.2 Forecast using Holt's method

Based on the trend observed in our data, we have utilized Holt's method for forecasting. For this purpose, we have chosen the smoothing constants α and β as 0.62 and 0.93, respectively. By applying these values to the data collected from 2016 to 2021, we have made a forecast for the year 2022, which we then compared to the actual value for 2022. Additionally, we have used this method to forecast the value for the year 2023.

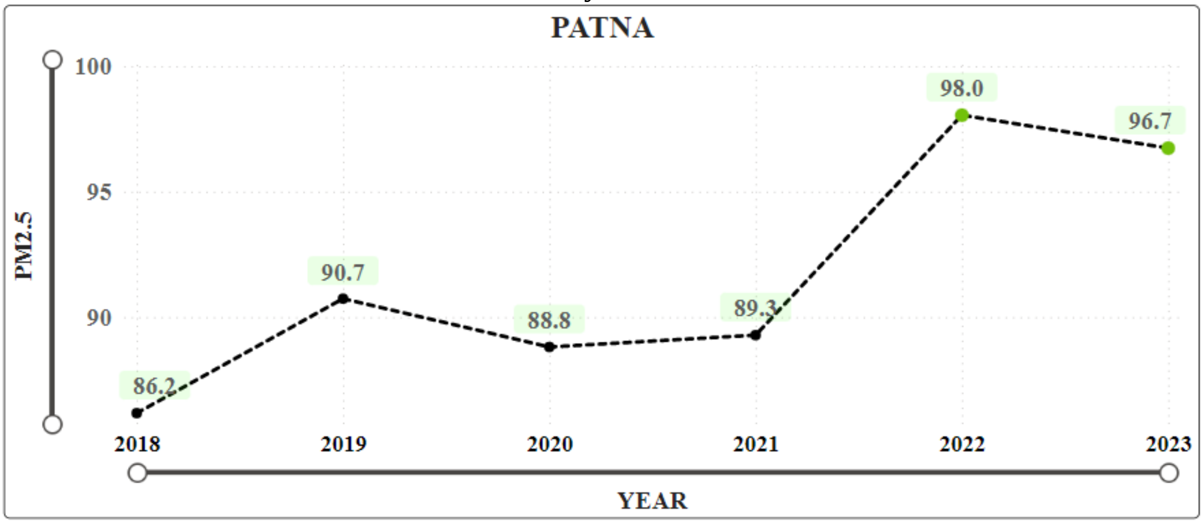


Figure 32 : Forecasting by Holt's method for Patna station

The predicted value we got from Holt's method is

Year	Actual	Predicted
2022	97.9528	98.04935488
2023		96.73497033

Table 25 : Prediction For Year 2022-23 By Holt's Method (Patna)

4.6.2.3 Forecast using Box-Jenkins procedure

In the past, we used classical methods and smoothing techniques to forecast the value of the parameter PM2.5 for the years 2022 and 2023 using data from 2016 to 2021. However, we have since utilized the Box-Jenkins procedure to determine the best-fitted model for our data from 2016 to 2022.
Using this model, we have forecasted the value of PM2.5 for the year 2023. Therefore, our prediction for the PM2.5 value in 2023 is based on the Box-Jenkins model applied to the data from 2016 to 2022. This approach offers a more sophisticated and accurate means of forecasting the value of the parameter, as it takes into account the complex relationships and patterns present in the data.

Test for stationarity

We have used augmented Dickey-Fuller test to check the stationarity.
Here the null hypothesis is the data is non-stationary and the alternative hypothesis is the data is stationary

Augmented Dickey-Fuller Test

Null: non-stationary
Alternative: stationary

Lag	ADF	P-value
1	-0.56634	0.9695

Table 26 : Augmented Dickey-Fuller Value & P-Value (Patna)

Initially, we performed the ADF (Augmented Dickey-Fuller) test on our data and found that the p-value for lag 1 was greater than 0.05, which means that we could not reject the null hypothesis at 5% level of significance based on our data. This indicated that our data was non-stationary.

To address this issue, we took the first-order difference of the data and performed the ADF test again. The result of the first difference data showed that the p-value for the lag 1 was now less than 0.05, indicating that we could reject the null hypothesis at 5% level of significance. This suggested that the first-order differenced data was stationary. Therefore, we can conclude that taking the first-order difference of the data made it stationary and suitable for further analysis & the model for our data will be ARIMA with d=1.

Model identification

As we have already discussed that, for model identification, we use ACF and PACF plots. These figures shows the ACF and PACF plot for station 2

Figure 33 : ACF & PACF of air pollutant Parameter PM2.5 for the Patna station

We can see that the above ACF plot shows a damped sinusoidal pattern but has no cut points. Here we can see that the above PACF plot has no cut points so, the suitable model may as ARIMA(0,1,1) Or ARIMA(1,1,0) model.
After considering suitable models for the station 2, we followed the procedure of Box-Jenkins' for each of the model one by one.

Parameter estimation of ARIMA(1,1,0) & ARIMA (0,1,1) model

MODELS	LOG L	φ1	θ1	θ2	σ2	K	AIC	BIC
ARIMA(0,1,1)	-24.26		-0.1679		189.1	2	52.52	52.41182
ARIMA(1,1,0)	**-24.08**	**-0.4428**			**173**	**2**	**52.16**	**52.05182**

Table 27 : Parameter Estimation

BIC of ARIMA(p,d,q) model
We use Bayesian Information Criterion to select one of two models. Here for ARIMA(1,1,0) we got a BIC score of 52.05182 which is less than the BIC score i.e.52.0518 for ARIMA(0,1,1) model.

Box-Pierce test

The result of Box-Pierce test for the model ARIMA(1,1,0) is shown below
X-squared = 0.4464, df = 1, p-value = 0.5041
From the score based approach we can see that p value is greater than 0.05 which means that the residuals are randomly distributed which means we can use this model for forecasting.

Forecasting using ARIMA(1,1,0) model

Here we have forecasting air pollutant parameter PM2.5 value for the year 2023 using data from 2016-2022 with the model ARIMA(1,1,0)

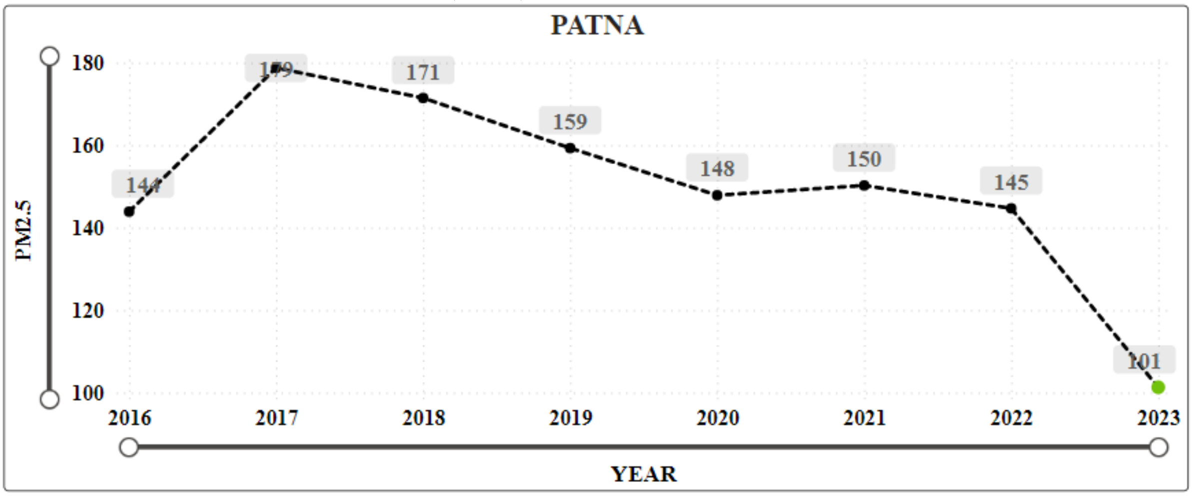


Figure 34 : Forecasting using ARIMA(1,1,0) model for Patna station

Following is the forecast value

Year	Predicted Value	Standard Error
2023	101.3132	13.15244

Table 28 : Forecast Value

4.6.3 Analysis of data for station 3 (Muzaffarpur)

4.6.3.1 Trend (Curve Fitting Method)

In order to analyse & forecast the average value of parameter for the station Patna we have used linear trend & fitted trend line which is shown in the given figure.

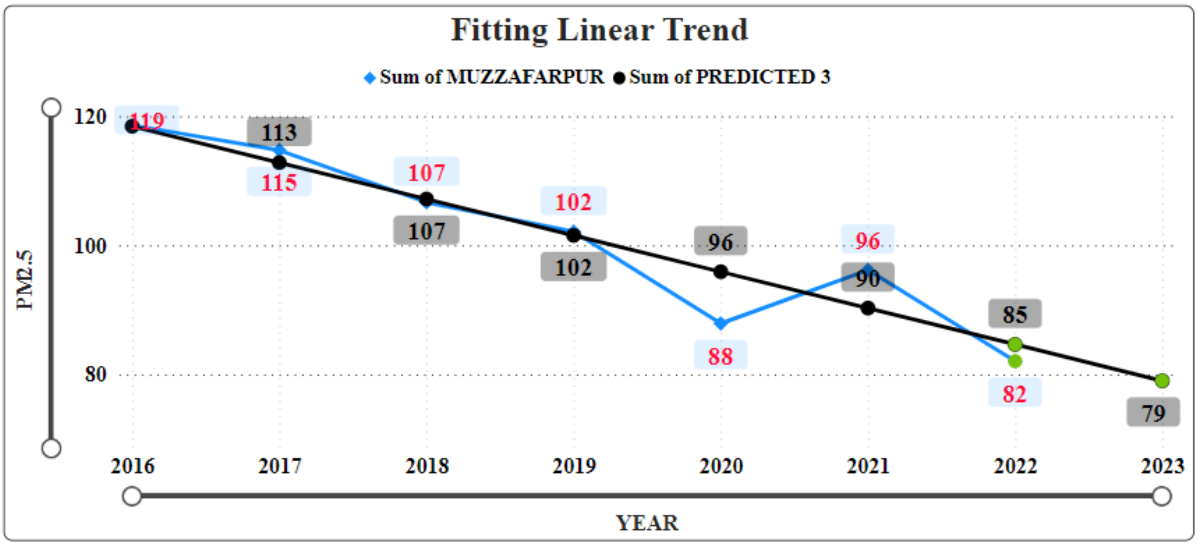


Figure 35 : Fitting Logarithmic curve on parameter PM2.5 for the station Muzzafarpur

The above graph shows a logarithmic trend curve fitted on the average value of the parameter PM2.5 for the years 2016-2021, which was collected from the central pollution board site. This curve was used to forecast the value for the year 2022, which was then compared to the actual value of 2022. Furthermore, a predicted value for the year 2023 has also been generated using this logarithmic trend curve.

Year	Actual	Predicted
2022	82.05559582	84.6457
2023		79.0108

Table 29 : Prediction For Year 2022-23 By Power Trend (Muzzafarpur)

The equation we get for this linear trend is
y = -5.6349*x + 124.09 ; where x is year

4.6.3.2 Forecast using Holt's method

Based on the trend observed in our data, we have utilized Holt's method for forecasting. For this purpose, we have chosen the smoothing constants α and β as 0.3 and 0.8, respectively. By applying these values to the data collected from 2016 to 2021, we have made a forecast for the year 2022, which we then compared to the actual value for 2022. Additionally, we have used this method to forecast the value for the year 2023.

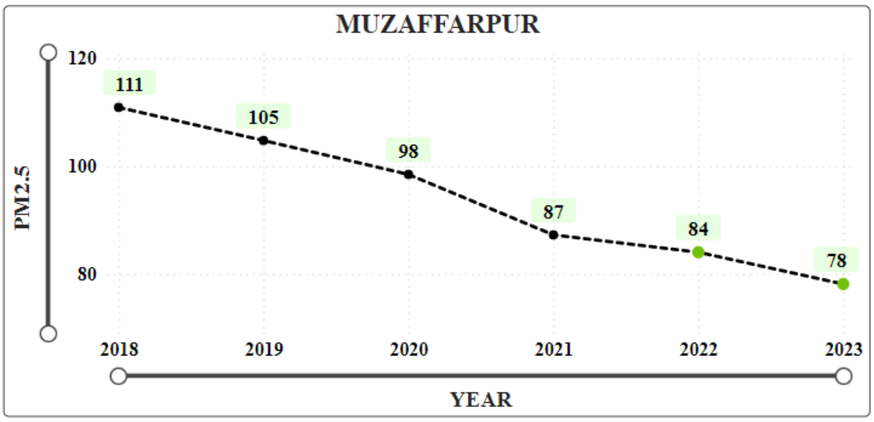


Figure 36 : Forecasting by Holt's method for Muzaffarpur station

The predicted value we got from Holt's method is

Year	Actual	Predicted
2022	82.05559582	84.038098
2023		78.146263

Table 30 : Prediction For Year 2022-23 By Holt's Method (Muzaffarpur)

4.6.3.3 Forecast using Box-Jenkins procedure

In the past, we used classical methods and smoothing techniques to forecast the value of the parameter PM2.5 for the years 2022 and 2023 using data from 2016 to 2021. However, we have since utilized the Box-Jenkins procedure to determine the best-fitted model for our data from 2016 to 2022.
Using this model, we have forecasted the value of PM2.5 for the year 2023. Therefore, our prediction for the PM2.5 value in 2023 is based on the Box-Jenkins model applied to the data from 2016 to 2022. This approach offers a more sophisticated and accurate means of forecasting the value of the parameter, as it takes into account the complex relationships and patterns present in the data.

Test for stationarity

We have used augmented Dickey-Fuller test to check the stationarity.
Here the null hypothesis is the data is non-stationary and the alternative hypothesis is the data is stationary

Augmented Dickey-Fuller Test

Null: non-stationary
Alternative: stationary

Lag	ADF	P-value
1	-1.086	0.9079

Table 31 : Augmented Dickey-Fuller Value & P-Value (Muzaffarpur)

Initially, we performed the ADF (Augmented Dickey-Fuller) test on our data and found that the p-value for lag 1 was greater than 0.05, which means that we could not reject the null hypothesis at 5% level of significance based on our data. This indicated that our data was non-stationary.

To address this issue, we took the first-order difference of the data and performed the ADF test again. The result of the first difference data showed that the p-value for the lag 1 was now less than 0.05, indicating that we could reject the null hypothesis at 5% level of significance. This suggested that the first-order differenced data was stationary. Therefore, we can conclude that taking the first-order difference of the data made it stationary and suitable for further analysis & the model for our data will be ARIMA with d=1.

Model identification

As we have already discussed that, for model identification, we use ACF and PACF plots. These figures shows the ACF and PACF plot for station 3

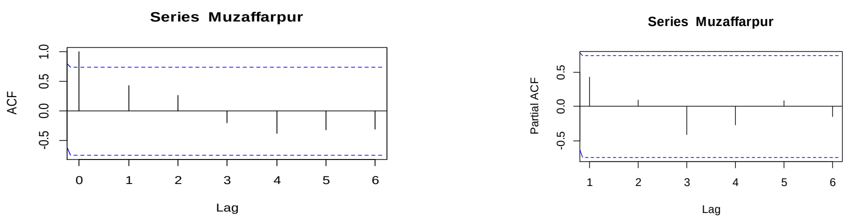


Figure 37: ACF & PACF of air pollutant Parameter PM2.5 for the Muzaffarpur station

We can see that the above ACF plot shows a damped sinusoidal pattern but has no cut points. Here we can see that the above PACF plot has no cut points so, the suitable model may as ARIMA(0,1,1) Or ARIMA(1,1,0) model.
After considering suitable models for the station 3, we followed the procedure of Box-Jenkins' for each of the model one by one.

Parameter estimation of ARIMA(1,1,0) & ARIMA (0,1,1) model

MODELS	LOG L	φ1	θ1	θ2	σ2	K	AIC	BIC
ARIMA(0,1,1)	-22.13		-0.0962		93.43	2	48.26	48.15182
ARIMA(1,1,0)	**-22.06**	**-0.2261**			**90.56**	**2**	**48.12**	**48.01182**

Table 32 : Parameter Estimation

BIC of ARIMA(p,d,q) model

We use Bayesian Information Criterion to select one of two models. Here for ARIMA(1,1,0) we got a BIC score of 48.0118 which is less than the BIC score i.e.48.1518 for ARIMA(0,1,1) model.

Box-Pierce test

The result of Box-Pierce test for the model ARIMA(1,1,0) is shown below
X-squared = 0.10909, df = 1, p-value = 0.7412
From the score based approach we can see that p value is greater than 0.05 which means that the residuals are randomly distributed which means we can use this model for forecasting.

Forecasting using ARIMA(1,1,0) model

Here we have forecasting air pollutant parameter PM2.5 value for the year 2023 using data from 2016-2022 with the model ARIMA(1,1,0)

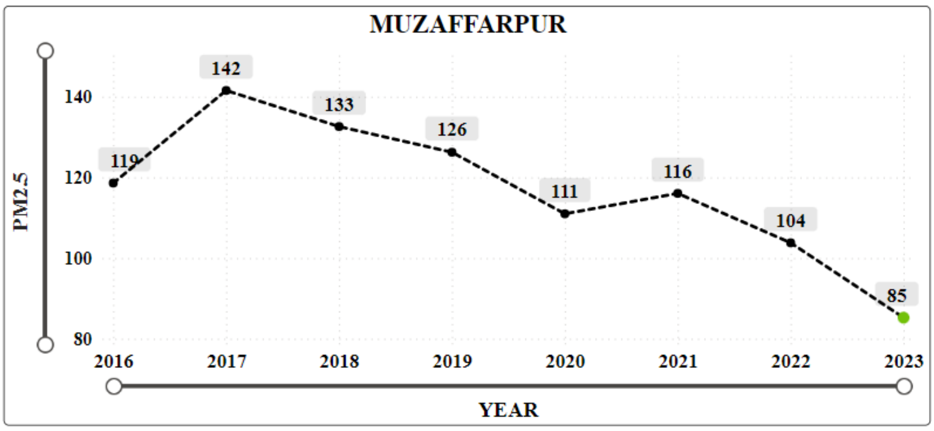


Figure 38 : Forecasting using ARIMA(1,1,0) model for Muzaffarpur station

Following is the forecast value

Year	Predicted Value	Standard Error
2023	85.25	9.516207

Table 33 : Forecast Value

4.7 Result & conclusion

In this project, we utilized PM2.5 data from five air quality monitoring stations in Bihar with the worst air quality index, namely Siwan, Muzaffarpur, Patna, Gaya, and Buxar. However, due to the unavailability of long-term data for two stations , we could only collect data from three stations (Gaya, Patna, and Muzaffarpur) for forecasting purposes. All data measures are in µg/m3 units and were obtained from the official website of the Central Pollution Control Board. Here we have considered the average value of 12 months for the parameter PM2.5 as it is the major affected parameter out of all other remaining parameters.

Year	Gaya	Patna	Muzaffarpur
2016	136.1753	143.9009	118.6105271
2017	205.2909	115.0376	114.7412894
2018	95.83572	120.5135	106.676888
2019	83.29376	105.9423	102.1276597
2020	61.18022	100.9611	87.89810067
2021	61.1827	105.5418	96.18192426
2022	67.11656	97.95287	82.05559582

Table 34 : Year-wise average value of parameter PM2.5 for all three stations

Based on the table provided, it can be inferred that the PM2.5 values from all three stations are within the moderate to very poor range according to the national air quality standards. Although this conclusion is not entirely conclusive, it does suggest that it may be possible to make interpretations about the air quality in Bihar in terms of the air pollutant PM2.5.
The main objective of our project is to forecast the value of the air pollutant PM2.5 using average data. To achieve this objective, we have forecasted the values for the years 2022 and 2023 by utilizing data from 2016 to 2021. For this purpose, we have employed three methods, which are as follows:

- Firstly, we estimated trends using power, logarithmic, and linear trend curves, and used these to forecast the average PM2.5 values.
- Secondly, we utilized Holt's method to forecast the average value of the air pollutant PM2.5.
- Finally, we applied the Box-Jenkins' procedure to identify a suitable model for our data and used this model to forecast the values of PM2.5.

The forecasted values of all the stations for the year 2022 with actual true average value is given in this table.

		Predicted Value		
Stations	**Actual**	**Year**	**Trend**	**Holt's Method**
Gaya	67.116556	2022	59.2394	67.1496
Patna	97.9528737	2022	96.7051	98.04935488
Muzaffarpur	82.0555958	2022	84.6457	84.038098

Table 35 : station wise actual & predicted value for the year 2022

- The "Actual" value of PM2.5 in 2022 at the Gaya station was 67.116 micrograms per cubic meter (μg/m3), which is a relatively high level of PM2.5.
- The "Trend" column indicates that the overall trend of PM2.5 levels at the Gaya station has been decreasing over time. This could be due to various factors, such as changes in local emissions, weather patterns, or environmental regulations.
- The "Actual" value of PM2.5 in 2022 at the Patna station was 97.9528737 micrograms per cubic meter (μg/m3), which is a relatively high level of PM2.5. This suggests that there may be air quality issues in the area that could have negative impacts on human health.
- The "Trend" column indicates that the overall trend of PM2.5 levels at the Patna station has been increasing over time. This could be due to various factors, such as changes in local emissions, weather patterns, or environmental regulations.
- The "Actual" value of PM2.5 in 2022 at the Muzaffarpur station was 82.0555958 micrograms per cubic meter (μg/m3), which is also a relatively high level of PM2.5. This suggests that there may be air quality issues in the area that could have negative impacts on human health.
- The "Trend" column indicates that the overall trend of PM2.5 levels at the Muzaffarpur station has been increasing slightly over time, with a predicted value of 84.6457 (μg/m3) in 2022. However, this trend is expected to reverse in 2023, with a predicted value of 79.0108 (μg/m3). This could be due to various factors, such as changes in local emissions, weather patterns, or environmental regulations.

The forecasted values of all the stations, for the year 2023, by all the three methods is given in this table.

		Predicted Value					
Stations	**Year**	**Trend**	**Holt's Method**	**Model Fitting**			
				Value	Model	AIC	BIC
Gaya	2023	54.8172	69.7031	64.85306	ARIMA(1,1,0)	68.04	67.93182
Patna	2023	93.7794	96.73497033	101.3132	ARIMA(1,1,0)	52.16	52.05182
Muzaffarpur	2023	79.0108	78.146263	85.25	ARIMA(1,1,0)	48.12	48.01182

Table 36 :Station Wise Predicted Value For The Year 2023

- Based on the table provided, a set of predicted values for three weather stations (Gaya, Patna, and Muzaffarpur) using an ARIMA(1,1,0) model for each station are calculated. The AIC and BIC values are also provided for each model, which are measures of the relative quality of the model fit, with lower values indicating better fit.
- The "Holt's Method" and "Model Fitting" values provide different predictions for future PM2.5 levels at the Gaya station. Holt's Method predicts a slight increase in PM2.5 levels from 67.1496 (μg/m3) in 2022 to 69.7031 μg/m3 in 2023, while the Model Fitting value predicts a slight decrease in PM2.5 levels to 64.85306 (μg/m3) in 2023. It's important to note that these are just predictions and may not necessarily reflect actual PM2.5 levels in the future.

- The "Holt's Method" and "Model Fitting" values provide different predictions for future PM2.5 levels at the Patna station. Holt's Method predicts a slight increase in PM2.5 levels from 98.04935488 (µg/m3) in 2022 to 96.73497033 (µg/m3)in 2023, while the Model Fitting value predicts a more significant increase in PM2.5 levels to 101.3132 (µg/m3) in 2023. It's important to note that these are just predictions and may not necessarily reflect actual PM2.5 levels in the future.
- The "Holt's Method" and "Model Fitting" values provide different predictions for future PM2.5 levels at the Muzaffarpur station. Holt's Method predicts a decrease in PM2.5 levels from 84.038098 (µg/m3) in 2022 to 78.146263 (µg/m3) in 2023, while the Model Fitting value predicts an increase in PM2.5 levels to 85.25 (µg/m3) in 2023. It's important to note that these are just predictions and may not necessarily reflect actual PM2.5 levels in the future.

4.8 Key Takeaways

The project utilized PM2.5 data from five air quality monitoring stations in Bihar with the worst air quality index, namely Siwan, Muzaffarpur, Patna, Gaya, and Buxar. However, only data from three stations (Gaya, Patna, and Muzaffarpur) were available for forecasting purposes due to the unavailability of long-term data for two stations. The average value of 12 months for the parameter PM2.5 was considered as it is the major affected parameter out of all other remaining parameters. The project's main objective was to forecast the value of the air pollutant PM2.5 using average data. To achieve this objective, the project utilized three methods: trend estimation using power, logarithmic, and linear trend curves, Holt's method, and the Box-Jenkins' procedure to identify a suitable model for the data. The values predicted by all the three methods for the year 2022 with the actual true average value were given in a table. The actual PM2.5 values for all three stations in 2022 were relatively high, suggesting air quality issues in the areas that could have negative impacts on human health. The overall trend of PM2.5 levels at the Gaya station has been decreasing over time, while the overall trend of PM2.5 levels at the Patna station has been increasing over time. The overall trend of PM2.5 levels at the Muzaffarpur station has been increasing slightly over time, with a predicted value of 84.6457 (µg/m3) in 2022. However, this trend is expected to reverse in 2023, with a predicted value of 79.0108 (µg/m3). The Holt's Method and Model Fitting methods provided different predictions for future PM2.5 levels at each station.

CHAPTER 5 : EXPLORING THE INTERSECTION OF PUBLIC AWARENESS, PERCEPTIONS, AND AIR POLLUTION IN GAYA CITY

5.1 Background

Gaya is a historic city located in the Indian state of Bihar. The city has a rich cultural heritage, including significant Buddhist and Hindu religious sites, and attracts a large number of tourists every year. However, like many other cities in India, Gaya is facing several environmental and developmental challenges that threaten the well-being of its residents and the sustainability of its future.

One of the most significant challenges facing Gaya is air pollution, which is caused by a variety of factors, including industrial emissions, vehicular traffic, and construction activities. The high levels of air pollution in the city have been linked to a range of health problems, including respiratory diseases, cardiovascular diseases, and cancer.

Another significant challenge facing Gaya is the loss of green areas, which has contributed to the degradation of the city's environment and quality of life. The loss of green areas has also led to a decline in biodiversity and ecosystem services, which are essential for maintaining the ecological balance of the city.

In addition to these environmental challenges, Gaya also faces a range of social and economic challenges, including poverty, unemployment, and inadequate infrastructure. These challenges have contributed to the overall low quality of life in the city and have hindered its development. Addressing these challenges will require a multi-faceted approach that involves the participation of government, civil society, and the private sector. Key interventions may include implementing pollution control measures, promoting sustainable development practices, enhancing the city's green spaces and biodiversity, improving infrastructure and services, and increasing awareness and education on environmental issues among the population.

5.2 Study variables

The study aimed to assess the level of awareness and perceptions about air pollution among the people of Gaya city. To achieve this objective, we have used several variables that were categorized into different groups. The first category of variables pertained to awareness about air pollution and included questions related to people's knowledge of air pollution, the resources they used for cooking, the ill effects of smoking on human health, and their use of masks outside their houses.

The second category of variables pertained to waste burning practices and included questions related to people's domestic waste burning habits and the effects of burning waste on the environment. The third category of variables focused on the seasonal variation of air pollution, and participants were asked about the season during which air pollution was most severe.

The fourth category of variables was related to the air quality of Gaya city, and participants were asked about their perception of the air quality in the city. The fifth category of variables pertained to the main causes of air pollution in Gaya city, and participants were asked to identify the primary sources of air pollution.

The sixth category of variables pertained to the ranking of types of pollution, and participants were asked to rank different types of pollution in order of severity. Finally, the seventh category of variables pertained to people's perceptions of the better place to live considering air pollution.

Overall, by analysing these variables, the researchers aimed to gain insight into the level of awareness, attitudes, and perceptions of people towards air pollution in Gaya city.

5.3 Study Objective

- To Examine Awareness Of Air Pollution Among People In Gaya City.
- To Gain Insight Into the Perceptions And Opinions That People Hold About The City Of Gaya.

5.4 Analysis

The quantitative data was analysed using IBM SPSS 22 and presented in tables. Initially, cross tabulation and chi-square tests for independence of attributes were performed between awareness and variables such as the ill effects of smoking, ill effects of burning domestic waste, smoking status, and mask usage, and corresponding p-values were reported in the table. Next, odds ratios were calculated between the dependent variable of awareness and each independent variable, such as place of living, adverse effects of smoking, and adverse effects of burning domestic waste, one at a time in accordance with the research objective. Finally, data visualization techniques such as pie charts and donut charts were utilized to gain insight into the perceptions held by people about Gaya.

5.5 Result

- **To Examine Awareness Of Air Pollution Among People In Gaya City.**

- Cross tabulation & chi-square test between dependent variable (awareness) & independent variable (mask usage, cooking resource, place of living etc.)

Table . cross tabulation & chi square test

		Awareness among people		
		Yes	No	P value
Do you use mask	Yes	119	2	0.079
	No	233	13	
Cooking resource	Biomass	28	5	0.003*
	Kerosene	6	0	
	LPG	318	10	
Place of living	Urban	330	13	0.277
	Rural	22	2	
adverse effects of smoking	Yes	328	11	0.021*
	No	24	4	
adverse effects of burning domestic waste	Yes	337	13	0.149

	No	15	2	

Key = *statistically significant (P-value<0.05)

Table 37 : Cross Tabulation & Chi Square Test

From the Table 37, it is evident that there is an association between awareness and cooking resource, as the p-value (0.003) is less than the significance level of 0.05. Additionally, we observed a significant association between knowing the adverse effects of smoking and awareness of air pollution, as the p-value was also less than 0.05

- Cross tabulation & odds ratio between awareness among people about air pollution & place of living, knowing (adverse effects of smoking & adverse effects of burning domestic waste).

		Awareness among people		
		Yes	No	odds
Place of living	Urban	330	13	1.082
	Rural	22	2	0.469
adverse effects of smoking	Yes	328	11	1.271
	No	24	4	0.256
adverse effects of burning domestic waste	Yes	337	13	1.105
	No	15	2	0.32

Table 38 : Cross Tabulation &Odds Ratio

Based on Table 38. we found that there were greater odds (1.082) of awareness in urban areas as compared to rural areas. Similarly, the odds (1.271) were greater for those who reported knowing the adverse effects of smoking, while the odds for not knowing the adverse effects of burning domestic waste were lower.

- **To Gain Insight Into the Perceptions And Opinions That People Hold About The City Of Gaya.**

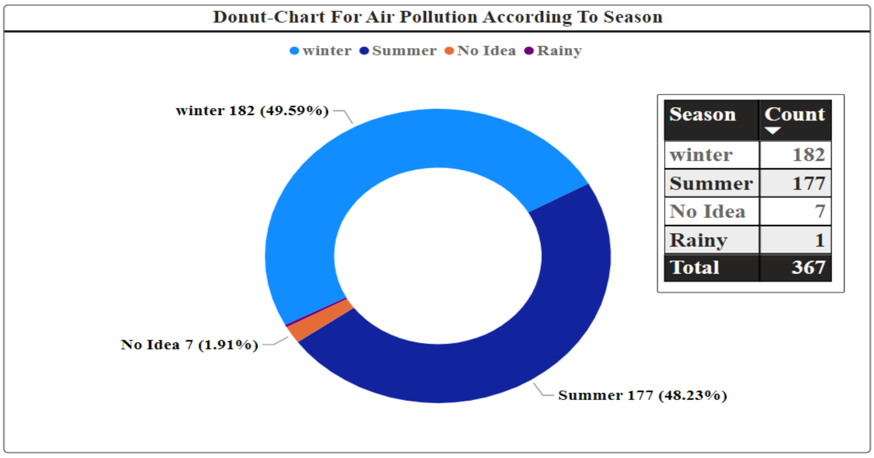


Figure 39 : Air pollution according to season.

As per Figure 39 the survey shows that a slightly higher percentage of respondents (49.59%) perceived air pollution to be high during the winter season, compared to 48.23% who found it to be high during the summer season.

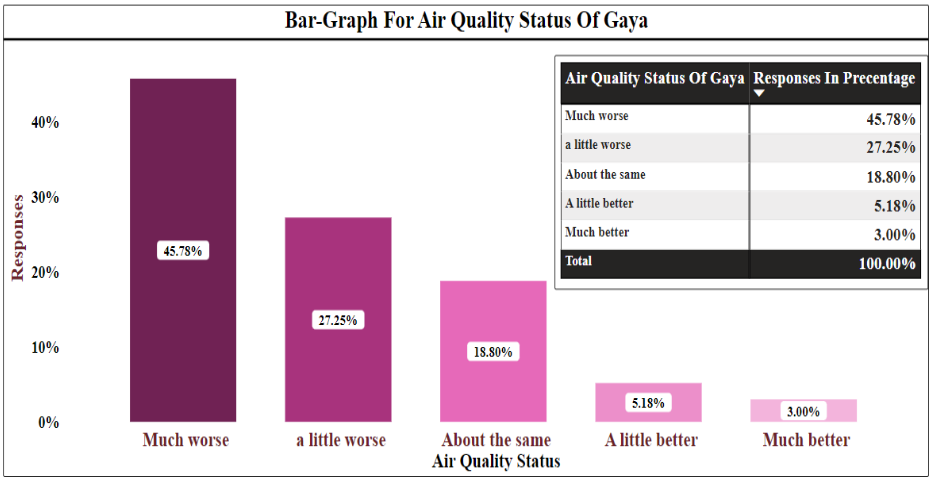


Figure 40 : Air quality status of Gaya

Figure 40. shows that according to the survey , a significant majority of the respondents (45.78%) rated the air quality of Gaya city as much worse, while only a small proportion (3%) found it to be much better.

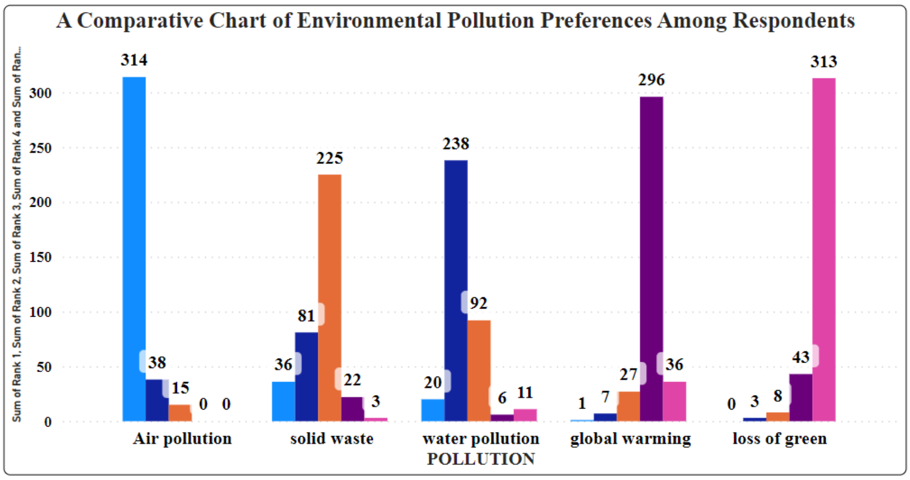


Figure 41 : Ranking of air pollution according to season

According to Figure 41. the majority of the respondents (314) view air pollution as the major environmental concern in Gaya, in contrast to the loss of green areas that causes relatively lesser environmental pollution.

Pollution	1	2	3	4	5	Weighted Score
Air pollution	314	38	15	0	0	1767
water pollution	20	238	92	6	11	1351
solid waste	36	81	225	22	3	1226
global warming	1	7	27	296	36	742
loss of green	0	3	8	43	313	435

Table 39 : Calculated weighted scores for types of pollution

The calculated weighted scores for the given factors indicate the perceived severity of each issue as reported by the respondents. According to the results, air pollution has the highest weighted score of 1767, suggesting that it is considered the most severe problem among the respondents. This is followed by water pollution with a weighted score of 1351, indicating that it is also perceived as a significant issue. The third-highest weighted score is for solid waste at 1226, which suggests that the respondents consider it to be a significant problem, but not as severe as air and water pollution.
In the paper "Assessing public preferences for urban green spaces conservation: A case study in the city of Valencia (Spain)" by Navarro-Hernández et al. (2012), the authors used a similar method to calculate the average score for each environmental attribute based on respondent rankings. They assigned weights of 5, 4, 3, 2, and 1 to the top five ranked attributes, respectively, and calculated the average weighted score for each attribute.

Here's the relevant excerpt from the paper:

"The importance of each attribute was estimated based on the ranking given by the respondents to each of them. Respondents ranked each attribute from 1 (most important) to 5 (least important). The importance was then calculated based on the sum of the weights assigned to each attribute by each respondent. In this case, the maximum possible weight was 5, and the minimum was 1. A higher score meant greater importance. Finally, the mean weighted score for each attribute was calculated."
Source: Navarro-Hernández, F. I., Córcoles-Jiménez, M. P., & Hernández-López, D. (2012). Assessing public preferences for urban green spaces conservation: A case study in the city of Valencia (Spain). Urban forestry & urban greening, 11(4), 383-388.

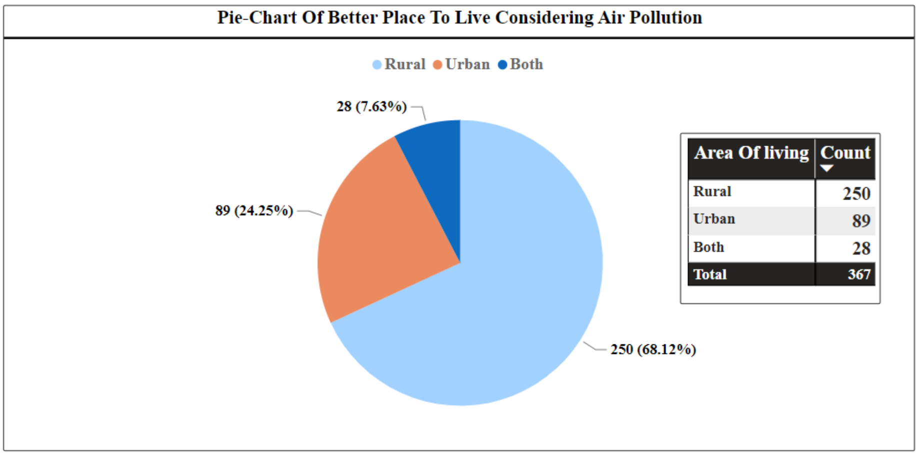


Figure 42 : Better place to live considering air pollution

Figure 42 says that according to the survey results, a significant majority of the respondents (68.12%) opined that rural areas offer a better living environment in terms of air pollution when compared to urban areas.

5.6 Discussion

The discussion highlights various environmental concerns and awareness in Gaya. The above findings suggest that being aware of issues such as cooking resources and the health risks of smoking is linked with environmental awareness. Urban residents and those who know the health effects of smoking are more likely to be environmentally conscious. Respondents perceive air pollution to be a significant issue in Gaya, with a higher percentage perceiving it to be high during winter than summer. The majority of respondents rated the air quality of Gaya city as much worse, indicating a pressing need to address air pollution. Additionally, air pollution is viewed as the major environmental concern in Gaya, followed by the loss of green areas. The weighted scores for different factors show that air pollution is considered the most severe problem by the respondents, followed by water pollution and solid waste. The findings suggest that addressing these issues can help improve the environment and public health in Gaya

CHAPTER 6 : A COMPREHENSIVE STUDY OF THE EFFECTS OF AIR POLLUTION ON HUMAN HEALTH

6.1 Background

Air pollution is a major environmental health issue that affects millions of people worldwide. Gaya, a city in India, is one of the many cities that faces significant air pollution problems due to industrialization, urbanization, and vehicular emissions. The city's rapid development in recent years has led to an increase in air pollution levels, posing a serious threat to the health and well-being of its residents. Exposure to high levels of air pollution has been linked to a range of health problems, including respiratory diseases, cardiovascular diseases, and even cancer. Studies have shown that air pollution can cause short-term effects such as irritation of the eyes, nose, and throat, as well as long-term effects such as reduced lung function and chronic respiratory diseases.

There is a need for a comprehensive study to examine the effects of air pollution on human health in Gaya. Such a study would provide a better understanding of the extent of the problem and its impact on the population. It would also help identify potential solutions and interventions to mitigate the negative health effects of air pollution. This chapter aims to provide an overview of the background and rationale for conducting a comprehensive study on the effects of air pollution on human health in Gaya.

6.2 Study Variables

Exposure to air pollution: This variable would measure the level of exposure to air pollution among the study population, which could be assessed using air quality monitoring data.

This study aims to investigate the impact of air pollution on the population of Gaya, with a particular focus on the factors that may influence exposure. The study will examine several key variables, including the type of address (urban or rural) of participants (variable 3), which may affect air pollution exposure due to differences in sources and distribution of pollutants. The study will also explore the status of using a mask outside the house (variable 29), which can provide valuable information on the level of awareness and protection measures taken by participants against air pollution exposure. By examining these variables, the study aims to provide insights into the impact of air pollution on residents in both urban and rural areas, as well as the use of protective measures against exposure. This information can be used to develop effective interventions and policies to reduce air pollution exposure and protect the health of the population in Gaya.

Health outcomes: This variable would measure the impact of air pollution on human health, which could include respiratory diseases, cardiovascular diseases, cancer, and other health outcomes.

This study aims to investigate the relationship between air pollution and symptoms and diseases in senior citizens. Key variables to be examined include whether seniors have consulted a doctor for air pollution-related symptoms (variable 23), and whether they have been diagnosed with air pollution-related diseases (variable 24). Symptoms related to air pollution (variable 25) and whether seniors have consulted a doctor for them (variable 26) will also be explored. Finally, the study will investigate if participants have been diagnosed with air pollution-related diseases (variable 27). The study seeks to provide valuable insights into the effects of air pollution on the health of senior citizens, and identify potential strategies to mitigate its negative impacts on this vulnerable population**.**

Demographic variables: This variable would include demographic characteristics such as age, gender, education, income, and occupation.

This study aims to investigate the relationship between air pollution in Gaya and factors related to exposure. Key variables to be examined include the working zone of participants (variable 11), time spent outdoors (variable 12), and duration of staying in Gaya (variable 8). By examining these variables, the study aims to provide valuable insights into the relationship between air pollution exposure and work, outdoor activities, and residency in Gaya. The findings can inform effective interventions to reduce air pollution exposure and mitigate negative health impacts on the population.

Lifestyle variables: This variable would include lifestyle factors such as smoking, physical activity, which could influence the susceptibility to the negative health effects of air pollution.

This study aims to investigate the impact of smoking habits on the exposure of individuals to air pollution in Gaya. The study will focus on several key variables, including the smoking status of participants (variable 13), which will help determine the extent to which smoking contributes to air pollution exposure. The study will also explore the smoking preferences of participants (variable 15), as well as the place of smoking (variable 17) and frequency of smoking in a day (variable 18), as these factors may influence air pollution exposure. By examining these variables, the study aims to provide insights into the impact of smoking habits on air pollution exposure, and identify potential strategies to reduce exposure in individuals who smoke. This information can be used to develop effective interventions and policies to reduce air pollution exposure and improve the overall health of the population in Gaya.

Environmental variables: This variable would include environmental factors such as temperature, humidity, and wind speed, which could affect the distribution and concentration of air pollutants.

The study variable for this research is the season in which air pollution is more prevalent, represented by (variable 30) . This variable aims to identify the specific times of the year when air pollution levels in Gaya are highest. By examining this variable, the study aims to provide valuable insights into the impact of seasonal changes on air pollution levels, and identify potential strategies to mitigate exposure during times of high pollution. The seasonality of air pollution is an important factor in understanding its impact on human health, as different seasons may bring different sources and levels of pollutants. By investigating this variable, the study can help inform the development of effective interventions and policies to reduce air pollution exposure and protect the health of the population in Gaya.

Spatial variables: This variable would include spatial factors such as the proximity to major roads, industries, and urban centers, which could influence the level of exposure to air pollution.

The study variables for this research are focused on the impact of air pollution on individuals in Gaya. (Variable 35) seeks to measure the extent to which air pollution affects individuals, while (variable 36) aims to identify the specific ways in which people are affected by air pollution. By examining these variables, the study aims to provide insights into the different impacts of air pollution on individuals, ranging from physical symptoms to broader effects on quality of life. This information can be used to develop effective interventions and policies to reduce air pollution exposure and mitigate its negative impact on human health in Gaya

Socioeconomic variables: This variable would include socioeconomic factors such as access to healthcare, housing conditions, and social support, which could affect the vulnerability to the negative health effects of air pollution.

This research aims to investigate the impact of household practices and air pollution on the population of Gaya. Variable 31 seeks to identify whether individuals burn their domestic waste, while variable 7 investigates the use of cooking practices and fuels, which can also affect air quality. Additionally, variable 35 is focused on measuring the extent to which air pollution is affecting individuals, ranging from physical symptoms to broader effects on quality of life. By examining these variables, the study aims to provide valuable insights into the specific household practices that contribute to air pollution, as well as the overall impact of air pollution on the population. This information can be used to develop effective interventions and policies to reduce air pollution exposure and improve the overall health of the population in Gaya. Understanding the extent to which air pollution is affecting individuals and households is a crucial step in addressing this critical public health issue.
These variables would help provide a comprehensive understanding of the effects of air pollution on human health in Gaya and inform the development of effective interventions to reduce the negative health impact of air pollution.

6.3 Study Objective

- To Determine Effect Of Air Pollution On Human Health
- To Determine Which Age Group Is Mostly Affected By Air Pollution.
- To Determine Which Disease Is Mostly Significant Among People In Gaya Due To Air Pollution.

6.4 Analysis

The quantitative data was analysed using IBM SPSS 22 and presented in tables as percentages and frequencies. Cross-tabulation and risk ratio analysis were conducted to explore the impact of air pollution on human health, based on various socio-demographic variables. Binary logistic regression was also performed to investigate the relationship between air pollution and human health, taking into account the socio-demographic variables.
The results were presented in the form of odds ratios, which showed the strength and direction of the relationships between the variables. Graphical representation was used to enhance the understanding of the data. The statistical methods used in this analysis included cross-tabulation, risk ratio analysis, and binary logistic regression. These methods provided a comprehensive analysis of the data and helped draw meaningful conclusions about the effects of air pollution on human health.

6.5 Result

- **To Determine Effect Of Air Pollution On Human Health**

Table No. We conducted a cross-tabulation and odds ratio analysis to examine the relationship between people affected by air pollution and their place of work (outdoor and indoor) and place of living (urban and rural).

		Affected by Air Pollution			
		Yes	No	P value	OR(CI)
Most of time at work	Outdoor	194	69		1.219(1.038,1.432)
	Indoor	62	42	0.008*	0.640(0.460,0.883)
currently staying	Urban	244	99		1.069(0.996,1.146)
	Rural	12	12	0.029*	0.434(0.201,0.935)
Smoking	Yes	108	30	0.004*	1.561(1.114,2.188)
	No	148	81		0.792(0.679,0.924)

Key = *statistically significant (P-value<0.05)

Table 40 : Cross Tabulation & Odds Ratio

- According to the Table 40, there are greater odds for individuals who work outdoors compared to those who work indoors, greater odds for individuals who reside in urban areas compared to those who reside in rural areas, and greater odds for individuals who smoke compared to those who don't smoke.
- Based on the table, p-value for place of work,place of residence & those who don't smoke are resp.(0.0008,0.029,0.004) which are less than α level of significance. Hence, we may conclude that there is an association between individuals affected by air pollution and their place of work, place of residence, and smoking habits

We Conducted cross-tabulation and binary logistic regression with odds ratio analysis to examine the relationship between individuals affected by air pollution and the zone they are living in.

		Affected by Air Pollution			
		Yes	No	P value	OR(CI)
Zone	Zone A	105	30		2.242 (1.215,3.942)
	Zone B	87	40		1.393 (0.810,2.396)
	Zone C	64	41	0.018*	1

Key = *statistically significant (P-value<0.05)

Table 41 : Cross Tabulation & Binary Logistic Regression

- Based on the binary logistic regression analysis, individuals living in Zone A and Zone B are more likely to be affected by air pollution compared to those living in Zone C, which is the reference category. The odds ratio for Zone A is 2.242, indicating that individuals living in Zone A have 2.242 times greater odds of being affected by air pollution compared to those in Zone C. Similarly, the odds ratio for Zone B is 1.393, indicating that individuals living in Zone B have 1.393 times greater odds of being affected by air pollution compared to those in Zone C. Therefore, it can be inferred that individuals living in Zone A and Zone B are at a higher risk of exposure to air pollution and may require targeted interventions to reduce their exposure and mitigate the negative health impacts of air pollution.

We conducted a cross-tabulation and odds ratio analysis to examine the relationship between respondent who got affected by any severe disease and their place of work (outdoor and indoor) and smoking status.

		Disease To respondent			
		Yes	No	P value	OR(CI)
Most of time at work	Outdoor	57	206	0.007*	1.239(1.092,1.405)
	Indoor	10	94		0.476(0.263,0.864)
Do You Smoke	Yes	43	95	0*	2.027(1.588,2.587)
	No	24	205		0.524(0.377,0.729)
Symptoms	Yes	64	3	0*	2.866(2.422,3.391)
	No	100	200		0.067(0.022,0.204)

Key = *statistically significant (P-value<0.05)

Table 42 : Cross Tabulation & Odds Ratio

- As per Table No 42. There are greater odds for those who work outdoor (1.239) as compare to those who work in indoor (0.476) there is more chance of having affected to the diseases by air pollution
- According to the analysis, smokers have significantly higher odds of developing health issues with odds ratio of 2.2027 as compared to non-smokers with odds ratio of 0.524.
- According to Table No.42, respondents who were exposed to air pollution and experienced symptoms had significantly greater odds of being affected by the disease with an odds ratio of 2.866, whereas those who did not experience symptoms had significantly lower odds of being affected with an odds ratio of 0.067.

We conducted a cross-tabulation and odds ratio analysis to examine the relationship between senior citizens (SC) who got affected by any severe disease and symptoms to which they are exposed and smoking status.

		Disease To SC			
		Yes	No	P value	OR(CI)
Symptoms	Yes	89	9	0*	16.269(8.619,30.708)
	No	4	144		0.046(0.018,0.119)
Do You Smoke	Yes	33	12	0*	4.524(2.462,8.313)
	No	60	141		0.700(0.598,0.820)

Key = *statistically significant (P-value<0.05)

Table 43 : Cross Tabulation & Odds Ratio

- "Based on Table No. 43, it was found that senior citizens who were exposed to air pollution and experienced symptoms had significantly higher odds of being affected by the disease, with an odds ratio of 16.269. Conversely, those who did not experience symptoms had significantly lower odds of being affected, with an odds ratio of 0.046."
- According to the analysis, smokers in senior citizens have significantly higher odds of

developing health Diseases with odds ratio of 4.524 as compared to non-smokers with odds ratio of 0.700.

▪ To Determine Which Age Group Is Mostly Affected By Air Pollution.

We conducted a cross-tabulation and Binary logistic regression utilizing odds ratio analysis to examine the relationship between Individuals affected by Air Pollution and the age group most impacted by it

		Affected by Air Pollution			
		Yes	No	P value	OR(CI)
AGE	15-30	99	51		1.387(0.419,4.587)
	31-45	105	29		2.586(.764,8.752)
	45-60	45	26		1.236(.356,4.294)
	60 Above	7	5	0.049*	1

Key = *statistically significant (P-value<0.05),1=Reference Category

Table 44 : Cross Tabulation & Binary Logistic Regression

- Based on the Table 44, the odds ratios indicate that individuals in the age groups of 31-45, 15-30, and 45-60 are more likely to be affected by air pollution compared to those in the reference group of 60 and above. Specifically, the odds ratio for the age group 31-45 is 2.586, indicating that this group is 2.586 times more likely to be affected by air pollution compared to the reference group. Similarly, the odds ratio for the age group 15-30 is 1.387, and for the age group 45-60 it is 1.236, indicating that these age groups are also at higher risk for being affected by air pollution. Overall, these findings suggest that younger and middle-aged individuals are more likely to be impacted by air pollution.

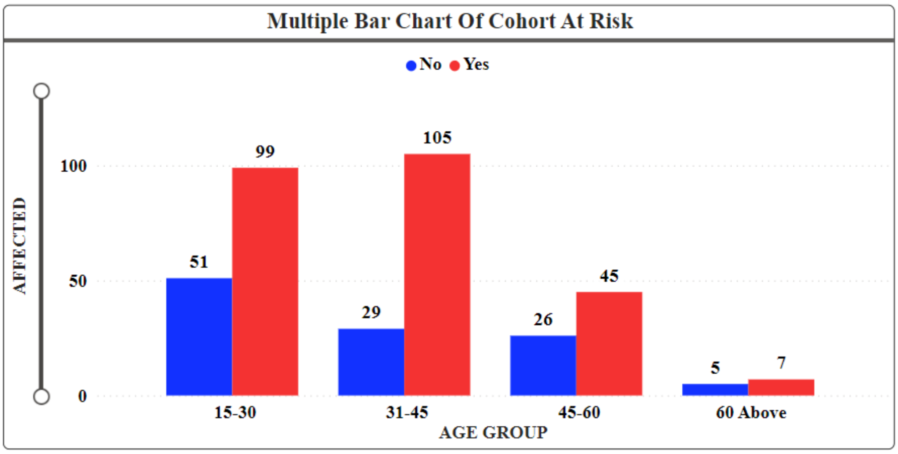


Figure 43 : Multiple Bar chart of cohort At Risk

The multiple chart of cohort at risk provides information about the number of individuals affected by air pollution in different age groups. The chart indicates that the highest number of individuals affected by air pollution is in the age group of 31-45, followed by the age group of

15-30. The age groups of 45-60 and 60 above also have a significant number of individuals affected by air pollution.
One possible explanation for this pattern could be that individuals in the age group of 31-45 and 15-30 may be more active and spend more time outdoors, which can increase their exposure to air pollution. Additionally, individuals in these age groups may be more likely to engage in behaviours that contribute to air pollution, such as driving cars or using certain household products.

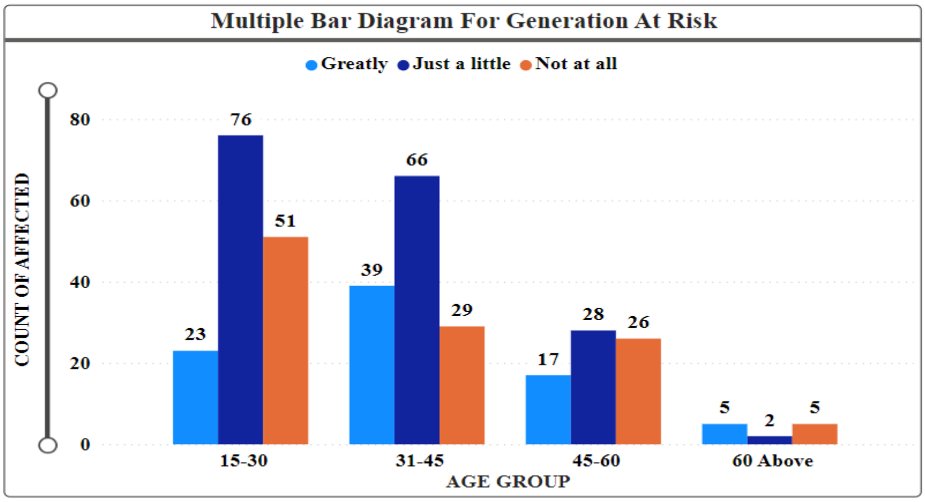


Figure 44 : Multiple Bar Diagram For Generation at Risk

The multiple bar chart for Generation At risk shows the distribution of individuals across different age groups who are greatly affected by air pollution, affected a little, and not affected at all. The age groups represented are 15-30, 31-45, 45-60, and 60 Above. The chart shows that the 31-45 age group has the highest proportion of individuals who are greatly affected by air pollution, followed by the 15-30 age group. The 45-60 age group and 60 Above have a lower chance of getting affected by air pollution.

To Determine Which Disease Is Mostly Significant Among People In Gaya Due To Air Pollution.

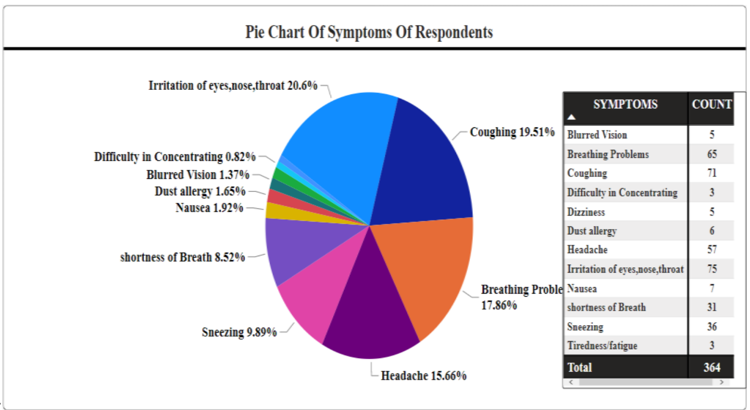


Figure 45 : Pie chart of Symptoms of Respondents

This pie chart displays the percentage of respondents experiencing different symptoms due to air pollution. The highest percentage of respondents experienced irritation of eyes, nose, and throat, accounting for 20.6% of the responses. Coughing was the second most common symptom reported by 19.51% of the respondents, followed by breathing problems (17.86%) and headache (15.66%). Sneezing was reported by 9.89% of the respondents, while shortness of breath was reported by 8.52% of them. Other symptoms reported included nausea, dust allergy, blurred vision, dizziness, difficulty in concentration, and tiredness. This chart provides insights into the types of symptoms that people in Gaya are experiencing due to air pollution, which can inform interventions and policies aimed at reducing air pollution exposure and improving public health.

To determine which disease is most significant among people in Gaya due to air pollution, we utilized graphical representation to show which diseases are more prevalent.

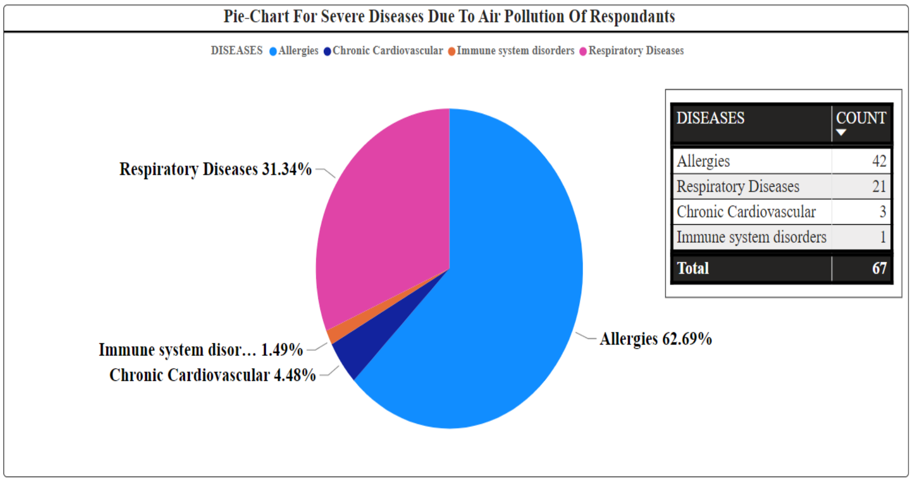


Figure 46 : Pie chart for severe Diseases Due to Air Pollution of Respondents

Based on the pie chart representing the prevalence of different types of diseases caused by air pollution in Gaya, allergies and respiratory diseases are the most prevalent, with 62.69% and 31.34% respectively. This is consistent with the symptoms reported by respondents, with 20.6% reporting irritation of eyes, nose, and throat, and 19.51% reporting coughing, both of which are common symptoms of allergies and respiratory diseases.

Furthermore, 17.86% of respondents reported breathing problems, which could be related to respiratory diseases. Headache, reported by 15.66% of respondents, is also a common symptom of allergies and respiratory diseases. Sneezing, reported by 9.89% of respondents, is a common symptom of allergies, and shortness of breath, reported by 8.52% of respondents, is a common symptom of respiratory diseases.

Overall, there appears to be a strong correlation between the prevalence of diseases caused by air pollution and the reported symptoms of the respondents in Gaya. This suggests that efforts should be made to address the sources of air pollution in the area, as it is having a significant impact on the health of the population.

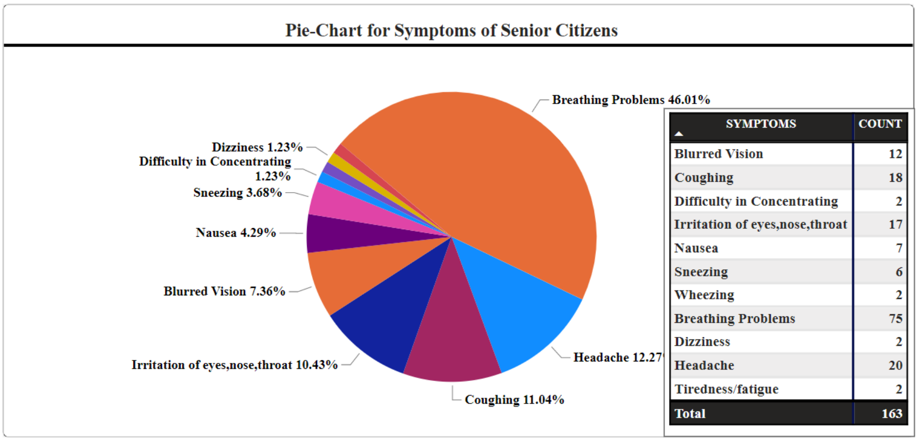


Figure 47 : Pie chart of Symptoms of Senior Citizens

This pie chart displays the prevalence of symptoms among senior citizens in Gaya, caused by air pollution. Breathing problems appear to be the most common symptom, with 46.01% of respondents reporting it, followed by headache (12.27%) and coughing (11.04%). Other symptoms reported include irritation of eyes, nose, and throat (10.43%) and blurred vision (7.36%). The chart provides an overview of the impact of air pollution on senior citizens in the area and highlights the need for action to improve air quality.

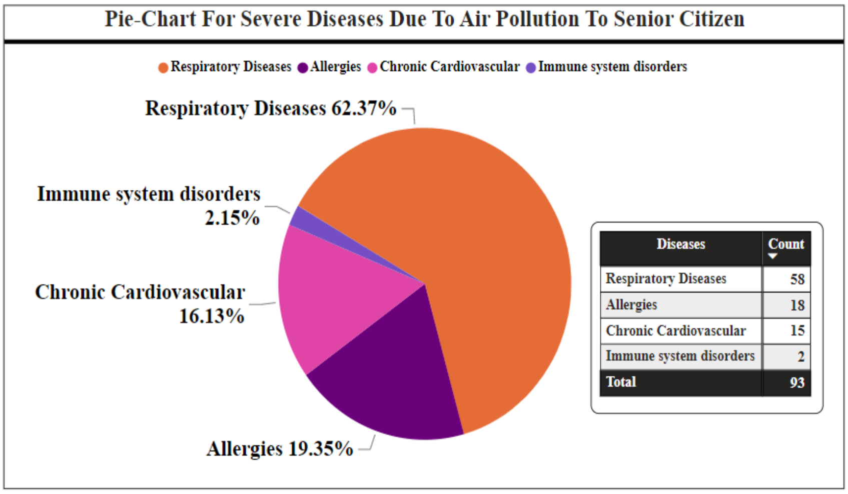


Figure 48 : Pie chart for severe Diseases Due to Air Pollution of Senior Citizens

This pie chart represents the distribution of severe diseases due to air pollution among senior citizens. The chart indicates that the majority of the senior citizens suffer from respiratory diseases, which account for 62.37% of the total cases. The second most significant disease is allergies, accounting for 19.35% of the cases. Chronic cardiovascular diseases and immune system disorders represent a smaller portion of the cases, with 11.83% and 6.45% respectively. The chart highlights the serious health impacts of air pollution on senior citizens, particularly regarding respiratory diseases and allergies.

Note:- The comparison of the two pie charts indicates a clear correlation between the prevalence of respiratory diseases and breathing problems among senior citizens in Gaya due to air pollution. This suggests that air pollution is a significant contributing factor to the development of respiratory diseases and related symptoms among the senior population. The high percentage of seniors suffering from respiratory diseases in the disease pie chart is reflected in the high prevalence of breathing problems reported in the symptom pie chart. Similarly, the prevalence of allergies in the disease pie chart is reflected in the high percentage of seniors reporting irritation of eyes, nose, and throat in the symptom pie chart.

6.6 Discussion:-

The various charts provide insights into the relationship between air pollution and its impact on different factors such as age, place of work and residence, smoking habits, and symptoms experienced by individuals.

The binary logistic regression analysis revealed that individuals living in Zone A and Zone B are more likely to be affected by air pollution compared to those living in Zone C, indicating the need for targeted interventions to mitigate the negative health impacts of air pollution on these populations. Additionally, individuals who work outdoors, reside in urban areas, and smoke are also at higher risk of being affected by air pollution.

The analysis also found that senior citizens, smokers, and individuals in the age groups of 31-45, 15-30, and 45-60 are more likely to be affected by air pollution, highlighting the need for targeted interventions to reduce their exposure.

The multiple bar charts and pie charts provide a visual representation of the distribution of individuals affected by air pollution across different age groups and the types of symptoms experienced by them. The charts show that younger and middle-aged individuals are more likely to be impacted by air pollution and the most commonly reported symptoms include irritation of eyes, nose, and throat, coughing, breathing problems, and headaches

CHAPTER 7: INVESTIGATING THE IMPACT OF AIR POLLUTION ON VARIOUS OCCUPATIONS IN GAYA

7.1 Background

Air pollution is a growing concern in many cities, including Gaya. It has been observed that different occupational groups are affected by air pollution to varying degrees. Therefore, it is important to identify which occupation is most affected by air pollution in Gaya. This information can help policymakers and stakeholders to prioritize interventions and develop targeted strategies to mitigate the impact of air pollution on the health and well-being of the population.

Hence, the objective of this study is to determine which occupation is mostly affected by air pollution in Gaya. By identifying the occupation that is most affected by air pollution, this study aims to contribute to a better understanding of the impact of air pollution on different occupational groups in the city. This, in turn, can inform the development of effective policies and interventions to improve air quality and protect the health of the population.

7.2 Study variables

In this study, we are interested in understanding the relationship between air pollution and occupation among the residents of Gaya. The two key variables in our investigation are (Var.35) and (Var.9).

(Var.35) represents the level of perceived impact of air pollution on the individuals. This variable has been measured on a scale ranging from "not at all affected" to "greatly affected". The response of each participant is important in determining the impact of air pollution on their health and wellbeing.

(Var.9) represents the occupation of each participant. This variable is significant in determining the impact of air pollution on different groups of people. It is widely known that certain occupations, such as construction workers or traffic police, are more exposed to air pollution due to the nature of their work. By considering the occupation of each participant, we aim to explore whether some groups of people are more affected by air pollution than others.

By examining the relationship between these two variables, we hope to gain insights into the impact of air pollution on the occupational health of residents in Gaya. Through our investigation, we aim to identify the occupations that are most affected by air pollution, which can inform policy and interventions to mitigate the impact of air pollution on the health and wellbeing of the people of Gaya.

7.3 Study objective

- **To Determine Which Occupation Is Mostly Affected By Air Pollution**

7.4 Analysis

The qualitative data was analysed using IBM SPSS 22 and presented in tables. Initially, cross tabulation and chi-square tests for independence of attributes were performed between people affected by air pollution in Gaya and their occupation and corresponding p-value was reported in the table. Next, y using multinomial logistic regression odds ratios were calculated between the dependent variable people affected by air pollution in three levels (greatly, just a little ,not affected at all)and independent variable their occupation in accordance with the research objective.

7.5 Result

We conducted Cross tabulation & chi-square test between people affected by air pollution & their occupation.

		35.To what extent is the air pollution affecting you ?			p-value
		Greatly	Just a little	Not at all	
occupation	Business	28	53	36	0.018
	Government job	4	7	6	
	Private job	23	39	23	
	Worker	24	32	15	
	Studying	3	31	27	
	Other	2	10	4	
Total		84	172	111	

Table 45 : Cross Tabulation & Chi-Square Test

Based on the information provided in the table, it can be observed that the p-value of 0.018 is less than the predetermined level of significance (α) of 0.05. This suggests that there is a statistically significant association between occupation and the extent to which respondents report being affected by air pollution. Therefore, we can conclude that there is a relationship between occupation and the reported impact of air pollution on individuals' health.

We Conducted The Odds ratio is calculated by multiple logistic regression between people affected by air pollution & their occupation.

To what extent air pollution is affecting particular occupation		Odds(CI)
Greatly	Business	1.556(0.226,9.112)
	Govt. Job	1.333(0.161,11.075)
	Private Job	2.000(0.333,12.016)
	Worker	3.200(0.521,19.668)
	Student	0.222(0.028,1.769)
	Other	1

Reference category = Not at all

Table 46 : Odds Ratio Is Calculated By Multiple Logistic Regression

From above table 46 we can say that, considering "not at all" as the reference category, people having occupation as "private job" are more likely to say that they are "greatly" affected by air pollution in comparison with the ones having field as "other" considered in this context. Similarly we can say that , considering "not at all" as the reference category, people having occupation as "worker" are more likely to say that they are "greatly" affected by air pollution in comparison with the ones having field as "other" considered in this context.

We conducted the relative percentage was computed for each occupation, considering the respondents greatly affected by air pollution in Gaya.

Count						
		35.To what extent is the air pollution affecting you ?				
		Greatly	Just a little	Not at all	Total	Percent of greatly affected by air pollution
occupation	Business	28	53	36	117	23.93162
	Government job	4	7	6	17	23.52941
	Private job	23	39	23	85	27.05882
	Worker	24	32	15	71	33.80282
	Studying	3	31	27	61	4.918033
	Other	2	10	4	16	12.5

Table 47 : Ralative Percentage

Based on the data presented in the table, it can be concluded that individuals in the occupation of "workers" are the most affected by air pollution. This is evident as the relative percentage (33.80282) of individuals greatly affected by air pollution in the "workers" category is higher than any other occupation.

7.6 Discussion

The analysis of the data revealed a statistically significant association between occupation and the extent to which respondents report being affected by air pollution. Specifically, individuals in the occupation of "workers" are the most affected by air pollution, as evident by the higher percentage of individuals greatly affected by air pollution in this category compared to any other occupation. Additionally, people with a private job are more likely to report being greatly affected by air pollution compared to those in other fields.

7.7 Conclusion

The research conducted in Gaya, India, highlighted various environmental concerns and the need for awareness about them. Respondents perceived air pollution to be the most significant issue, followed by the loss of green areas. Air pollution was considered the most severe problem, followed by water pollution and solid waste. The findings suggest that addressing these issues can help improve the environment and public health in Gaya.

The relationship between air pollution and its impact on different factors such as age, place of work and residence, smoking habits, and symptoms experienced by individuals was also studied. The analysis revealed that individuals living in Zone A and Zone B, those who work outdoors, reside in urban areas, and smoke are more likely to be affected by air pollution. Senior citizens and individuals in the age groups of 31-45, 15-30, and 45-60 are also at higher risk of being affected by air pollution. Younger and middle-aged individuals were found to be more likely to be impacted by air pollution, with symptoms including irritation of eyes, nose, and throat, coughing, breathing problems, and headaches.

The analysis of the data also revealed a statistically significant association between occupation and the extent to which respondents report being affected by air pollution. Individuals in the occupation of "workers" were found to be the most affected by air pollution, and people with a private job were more likely to report being greatly affected by air pollution compared to those in other fields.

Overall, the study highlights the urgent need for targeted interventions to mitigate the negative health impacts of air pollution, particularly in high-risk areas such as Zone A and Zone B, and among vulnerable groups such as senior citizens and workers. Addressing these issues can help improve the environment and public health in Gaya, India.

CHAPTER 8 : CONCLUSION

The project analyzed PM2.5 data from five air quality monitoring stations in Bihar with the worst air quality index. Three stations (Gaya, Patna, and Muzaffarpur) were used to forecast PM2.5 levels using trend estimation, Holt's method, and Box-Jenkins' procedure. The actual PM2.5 values for all three stations in 2022 were relatively high, indicating air quality issues in the areas that could negatively impact human health. The overall trend of PM2.5 levels at the Gaya station has been decreasing over time, while the overall trend of PM2.5 levels at the Patna station has been increasing over time. The Holt's method and Model Fitting methods provided different predictions for future PM2.5 levels at each station.

Now to analyze the effects of PM2.5 and to check the coincidence of the results from secondary data study we conducted a primary survey based study which gave the following conclusion.

The research conducted in Gaya highlighted environmental concerns and the need for awareness about them. Respondents perceived air pollution to be the most significant issue, followed by the loss of green areas. The analysis revealed that individuals living in Zone A and Zone B, those who work outdoors, reside in urban areas, and smoke are more likely to be affected by air pollution. Senior citizens and individuals in certain age groups are also at higher risk of being affected by air pollution. Individuals in the occupation of "workers" were found to be the most affected by air pollution. The study highlights the urgent need for targeted interventions to mitigate the negative health impacts of air pollution, particularly in high-risk areas and among vulnerable groups. Addressing these issues can help improve the environment and public health in Gaya, India.

ANNEXURE

8.1 Questionnaire :

Confidentiality Statement:
Data collected through this questionnaire shall be kept confidential and will be reported numerically only (without revealing individual identification)

Socio-demographic and Economic data		
Serial No.	Questions	Possible answer
1	Gender	1) Male 2) Female
2	What is your age	1) BELOW 15 2) 15-30 3) 31-45 4) 45-60 5) 60 Above
3	Where are you staying currently	1) Rural 2) Urban
4	What is your ward number	
5	In which locality do you spend most of your time in a day	1) Kareemgunj[Zone A] 2) Gaya Collectorate[Zone B] 3) Kushdheera[Zone c]
6	How many senior citizens are there in your house	1) 0 2) 1 3) 2 4) More than 2
7	If yes, do they smoke	1) Yes 2) No
8	Out of these, what do you use for cooking	1) LPG 2) Biomass 3) Kerosene
9	For how long have you been living in Gaya now	1) Less than 1 2) 1-5 3) 5-10 4) 10-15 5) More than 15
10	What is your current occupation	1) Business 2) Government job 3) private job 4) worker 5) studying
11	In which field are you working	1) Construction 2) Transport 3) Chemical Industry 4) Hawkers 5) Mining field 6) Other:
12	Where do you spend most of your time at work	1) Outdoor 2) Indoor

13	. How much time do you spend outside office/home(in hours)?	1) 0-3 2) 3-6 3) 6-9 4) 9-12 5) 12+
14	Do you smoke	1) Yes 2) No
15	If No, then in past have you ever smoked	1) Yes 2) No
16	If yes, then what do you prefer	1) Cigarette 2) Hookah 3) Bidi
17	Do you know the ill effects of smoking on human health	1) Yes 2) No
18	Where do you smoke	1) Inside home 2) Outside home
19	How frequently do you smoke in a day	1) Once 2) Twice 3) 3-4 times 4) More than 4 times
20	Do you have vehicle	1) Yes 2) No
21	What type of vehicle is it	1) CNG 2) Petrol 3) Diesel 4) Electric
22	Have you ever witnessed Black smoke emitted out of your vehicle	1) Yes 2) No
23	Check all the symptoms or discomfort Senior citizens are experiencing	1) No issues 2) Headache 3) Nausea 4) Dizziness 5) Tiredness/fatigue 6) Irritation of eyes,nose,throat 7) Breathing Problems 8) Coughing 9) Sneezing 10) Wheezing 11) Blurred Vision 12) Difficulty in Concentrating
24	Has he/she ever consulted doctor for the any of the above symptoms	1) Yes 2) No
25	Do the senior citizen has any of the health problem that may make symptoms worse	1) Respiratory Diseases 2) Chronic Cardiovascular 3) Immune system disorders 4) Allergies 5) No
26	Check all the symptoms or discomfort are you experiencing	1) No 2) Headache 3) Nausea

		4) Dizziness 5) Tiredness/fatigue 6) Irritation of eyes, nose, throat 7) Breathing Problems 8) Coughing 9) Sneezing 10) Wheezing 11) Shortness of Breath 12) Blurred Vision 13) Sinus Congestion 14) Difficulty in Concentrating
27	If yes, have you ever consulted doctor for the above symptoms	1) Yes 2) No
28	Do you have any of the health Conditions that may make symptoms worse	1) Respiratory Diseases 2) Chronic Cardiovascular 3) Immune system disorders 4) Allergies 5) No
29	. Are you aware of Air Pollution	1) Yes 2) No
30	Do you use Mask outside your house	1) Yes 2) No
31	According to you in which season do you think air pollution is the most?	1) Summer 2) winter 3) Rainy
32	Do you burn your Domestic waste	1) Yes 2) No
33	. Do you know the adverse effects of burning domestic waste	1) Yes 2) No
34	. How would you rate the overall air quality in your city	1) Much better 2) A little better 3) About the same 4) a little worse 5) Much worse
35	What do you think are the main causes of air pollution in your city?	1) Construction 2) Industrial sources/manufacturing facilities 3) Increasing use of air conditioner 4) Motor vehicles 5) Household cooking and heating 6) Population growth 7) Power plants 8) Smoke of Cigarettes 9) Waste disposal 10) Burning of Waste 11) Pollution from other Countries

36	Are you affected by air pollution	1) Yes 2) No
37	To what extent is the air pollution affecting you	1) Greatly 2) Just a little 3) Not at all
38	Please rank the following based on your level of concern [1-highest concern and 7- Lowest concern]	1) Air pollution 2) Water pollution 3) Garbage and solid waste 4) Global warming and climate change 5) Loss of green areas in city
39	According to you which area/region is better for living on considering air Pollution?.	1) Urban 2) Rural 3) Both
40	Do you have any suggestion to improve Air Quality of your City.	

8.2 Frequency

Variables	N (%)
Gender	
Female	5(1.4)
Male	362(98.6)
What is your age [in years]?	
15-30	150(40.9)
31-45	134(36.5)
45-60	71(19.3)
60 Above	12(3.3)
Where are you staying currently?	
Rural	24(6.5)
Urban	343(93.5)
Zone	
Zone A	135(36.8)
Zone B	127(34.6)
Zone C	105(28.6)
How many senior citizens are there in your house?	
0	121(33)
1	134(36.5)
2	99(27)
More than 2	13(3.5)
If YES, do they smoke?	
No	196(53.4)
Yes	45(12.3)
Not-Applicable	126(34.3)
Out of these, what do you use for cooking?	
Biomass	33(9)
Kerosene	6(1.6)
LPG	328(89.4)
For how long have you been living in Gaya now?	
Less than 1	4(1.1)
1-5	8(2.2)
5-10	29(7.9)
10-15	46(12.5)
More than 15	280(76.3)
What is your current occupation	
Business	117(131.9)
Government Job	17(4.6)
Private Job	85(23.2)
Worker	71(19.3)
Studying	61(16.6)

Other(lawyer, politician, etc.)	16(4.4)
Where do you spend most of your time at work?	
Indoor	104(28.3)
Outdoor	263(71.7)
How much time do you spend outside office/home(in hours)?	
0-3	36(9.8)
3-6	62(16.9)
6-9	88(24)
9-12	144(39.2)
12+	37(10.1)
Do you smoke	
No	229(62.4)
Yes	138(37.6)
If yes, then what do you prefer?	
Bidi	10(2.7)
Cigarette	125(34.1)
Hookah	3(.8)
Not applicable	229(62.4)
Do you know the ill effects of smoking on human health?	
No	28(7.6)
Yes	339(92.4)
Where do you smoke?	
Both	23(6.3)
Inside home	4(1.1)
Not applicable	229(62.4)
Outside home	111(30.2)
How frequently do you smoke in a day?	
3-4 times	33(9)
More than 4 times	7(1.9)
Not-Applicable	230(62.7)
Once	36(9.8)
Twice	61(16.6)
Do you have vehicle?	
No	94(25.6)
Yes	273(74.4)
What type of vehicle is it ?	
CNG	1(.3)
CNG, Petrol	2(.5)
Diesel	19(5.2)
Electric	14(3.8)
Not-Applicable	94(25.6)
Petrol	220(59.9)
Petrol, Diesel	16(4.4)

Petrol, Electric	1(.3)
Have you ever witnessed black smoke emitted out of your vehicle?	
Maybe	5(1.4)
No	235(64)
Not-Applicable	95(25.9)
Yes	32(8.7)
Has he/she ever consulted doctor for the any of the above symptoms?	
No	17(4.6)
Not applicable	269(73.3)
Yes	81(22.1)
Do the senior citizen has any of the health problem that may make symptoms worse	
Allergies	18(4.6)
Chronic Cardiovascular	15(4.1)
Immune system disorders	2(.5)
No	153(41.7)
Not applicable	121(33)
Respiratory Diseases	58(15.8)
If yes, have you ever consulted doctor for the above symptoms?	
No	86(23.4)
Not applicable	198(54.0)
Yes	83(22.6)
Do you have any of the health Conditions that may make symptoms worse ?	
Allergies	42(11.4)
Chronic Cardiovascular	3(.8)
Immune system disorders	1(.3)
No	300(81.7)
Respiratory Diseases	21(5.7)
Are you aware of Air Pollution?	
No	15(4.1)
Yes	352(95.9)
Do you use Mask outside your house?	
No	246(67)
Yes	121(32.4)
According to you in which season do you think air pollution is the most?	
No Idea	7(1.9)
Rainy	1(.3)
Summer	177(48.2)
Winter	182(49.6)
Do you burn your Domestic waste?	
No	320(87.2)
Yes	47(12.8)

Do you know the adverse effects of burning domestic waste ?	
No	17(4.6)
Yes	350(95.4)
How would you rate the overall air quality in your city?	
A little better	19(5.2)
A little worse	100(27.2)
About the same	69(18.8)
Much better	11(3)
Much worse	168(45.8)
Are You affected by air pollution	
Yes	256(69.8)
No	111(30.2)
To what extent is the air pollution affecting you ?	
Greatly	84(22.9)
Just a little	172(46.9)
Not at all	11(30.2)
According to you which area/region is better for living on considering air pollution?	
Both	29(7.9)
Rural	250(68.1)
Urban	88(24)

8.3 Reference

1. https://apps.who.int/gho/data/node.main.BODAMBIENTAIR?lang=en
2. https://www.aqi.in/blog/know-about-the-ill-effects-of-air-pollution-on-your-lungs/
3. https://www.aqi.in/blog/report-about-93-of-children-who-are-below-the-age-of-15-get-affected-by-toxic-air/
4. https://www.aqi.in/blog/what-is-particulate-matter-and-its-health-effects/
5. https://www.aqi.in/blog/new-who-air-quality-guidelines-2021/
6. Home | State of Global Air
7. https://www.aqi.in/blog/analysis-of-air-pollution-and-indian-cities/
8. https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ambient-air-pollution
9. https://aaqr.org/articles/aaqr-21-08-oa-0204
10. https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database/2018
11. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/intimate-partner-violence-lifetime
12. https://www.iqair.com/in-en/world-most-polluted-countries
13. https://urbanemissions.info/india-air-quality/india-satpm25/
14. https://www.cseindia.org/winter-pollution-levels-and-trends-point-to-a-national-air-quality-crisis-says-cse-s-latest-analysis11197#:~:text=Average%20winter%20pollution%20of%202021,all%20regions%20with%20some%20variation.
15. https://gcwgandhinagar.com/econtent/document/15876169600EVSAECC01_Air%20pollution.pdf
16. https://urbanemissions.info/blog-pieces/resources-energy-emissions-analysis-in-india/
17. https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health
18. Air Quality in India - Reanalyzed PM2.5 Pollution Data 1998-2020 - (urbanemissions.info)
19. Winter pollution levels and trends point to a national air quality crisis, says CSE's latest analysis (cseindia.org)
20. https://apps.who.int/gho/data/node.main.152?lang=en
21. World's Most Polluted Cities in 2022 - PM2.5 Ranking | IQAir
22. Upashe, S.P., Tekelab, T. & Mekonnen, J(2015). Assessment of knowledge and practice of menstrual hygiene among high school girls in Western Ethiopia. BMC Women's Health 15, 84 (https://doi.org/10.1186/s12905-015-0245-7).
23. Shahzadi et al. (2018) can be cited as the source of the statement that people must have knowledge about household waste disposal, and awareness among people about dealing with waste is essential. The study can be found at: https://doi.org/10.1016/j.heliyon.2018.e00766.
24. The study conducted by Kaza et al. (2018) can be cited as the source of the statement that cities with high levels of public awareness about waste management had lower waste generation rates and higher recycling rates. The study can be found at: https://doi.org/10.1016/j.resconrec.2017.06.028.

25. The study by Houghton et al. (2019) can be cited as the source of the statement that education and awareness programs on waste management can lead to a significant reduction in waste generation and an increase in recycling rates. The study can be found at: https://doi.org/10.1016/j.jclepro.2019.117622.
26. Khan, R., Chakraborty, D., & Kumar, P. (2021). Assessment of Municipal Solid Waste Management in Gaya City, Bihar, India. International Journal of Environmental Research and Public Health, 18(2), 674.
27. Bhagat, R. B., & Kumari, A. (2017). Assessment of Municipal Solid Waste Management Scenario of Gaya City (Bihar). International Journal of Environmental Science and Development, 8(1), 37-42.
28. Shahzadi, A., Hussain, I., & Abro, R. (2018). An Assessment of Knowledge, Attitude and Practices of Household Waste Disposal in Gaya, Bihar, India. Journal of Health and Pollution, 8(19), 180904.
29. Singh, S. K., Kumar, A., & Kumar, P. (2019). Municipal solid waste management in Gaya city, Bihar: A case study. Environmental Science and Pollution Research, 26(35), 35984-35997
30. Vijayaraghavan, K., Joshi, U. M., & Balasubramanian, R. (2015). Assessment of greenhouse gas emissions from different waste management practices. Journal of Cleaner Production, 96, 305-314.
31. Singh, R. P., Tyagi, V. V., & Pandey, S. K. (2018). Challenges and opportunities in urban waste management: A case study of Delhi city, India. Journal of Cleaner Production, 172, 3121-3131.
32. Community-Led Initiatives for Solid Waste Management in Low-Income Settlements: A Case Study of Mumbai, India" by Aditi Ramola and Suresh Jain (2019)
33. Solid Waste Management Practices of Households in a Suburban Area of Mumbai" by Shobha B. Shukla, Prachi Mahajan, and Ramesh Agarwal (2017)
34. Assessment of waste management practices and the role of local authorities in Hargeisa City, Somaliland" by Ali et al. (2020)
35. Assessment of municipal solid waste management practices and associated problems in Gondar town, Ethiopia" by Tadesse and Kumie (2014)
36. Awareness and practices of waste segregation at source: A case study of Chandigarh City, India" by Kumari et al. (2018)
37. Kumar, S., Kumar, S., & Singh, S. (2019). Assessment of Waste Management Practices and Challenges in Gaya City. International Journal of Environmental Sciences & Natural Resources, 16(2), 1-8.
38. Singh, A. K., & Gupta, P. (2019). Assessment of solid waste management practices in Gaya, India. Environmental Science and Pollution Research, 26(6), 5696-5706.
39. Oluwasola, A., Oyeyemi, O. T., & Odu, O. O. (2019). Health Implications of Poor Municipal Solid Waste Management Practices in Nigeria. Journal of Environmental and Public Health, 2019. https://doi.org/10.1155/2019/9207526
40. Focks, D. A., Daniels, E., Haile, D. G., & Keesling, J. E. (1995). Aedes aegypti (Diptera: Culicidae) movement influenced by availability of oviposition sites. Journal of Medical Entomology, 32(6), 706-710. https://doi.org/10.1093/jmedent/32.6.706
41. Carreon, M. A., Asa, A. D. T., & Lacson, J. M. (2013). Effectiveness of community-based solid waste management in reducing diarrheal incidence in rural Philippines. International Journal of Public Health and Epidemiology, 2(1), 7-14. http://www.openscienceonline.com/journal/archive2?journalId=717&issueId=791
42. Hygiene practices and solid waste management in low-income communities: a case study of Nairobi, Kenya" by Gakii and Gichaba (2019)

43. Assessment of hygiene practices and awareness among municipal solid waste workers in Delhi, India" by Jain et al. (2016)
44. Waste management and hygiene practices among households in urban slums of Mumbai, India" by Sarkar et al. (2013)
45. Hygiene practices and waste management in rural households of India: An exploratory study" by Suryavanshi et al. (2018)
46. Impact of hygiene education on school children's knowledge and practices regarding solid waste management in rural India" by Kavitha et al. (2018)
47. Akinbode, O.A. and Oluwande, P.A., 2020. Assessment of health implications of improper solid waste management in Nigeria. Heliyon, 6(11), p.e05311. https://doi.org/10.1016/j.heliyon.2020.e05311
48. Kansal, A., Garg, V. and Sandhu, C., 2013. Health hazards caused by improper waste management practices. Journal of community health, 38(4), pp.690-697. https://doi.org/10.1007/s10900-013-9706-8
49. Oliveras-Ferraros, C., Vrhovnik, P., Bianchi, F., Fernández-Baldor, Á. and Hernández-Castro, R., 2021. Household waste management in developing countries: The case of the Philippines. Journal of Cleaner Production, 290, p.125573. https://doi.org/10.1016/j.jclepro.2020.125573
50. Santos, C.A.M., Luiz, R.R., Costa, E.A., Otero, U.B., Galvão, L.A.C., Diniz, P.S. and Eloi-Santos, S.M., 2018. Effects of a municipal solid waste management program on the prevalence of dengue fever in an endemic region in Brazil. Waste Management, 80, pp.113-120. https://doi.org/10.1016/j.wasman.2018.07.022
51. Adedeji, A.A., Okareh, O.T. and Adetunde, A.I., 2020. Review of health impacts of poor waste management practices. International Journal of Sustainable Development & World Ecology, 27(5), pp.440-450.
52. Akinyele, A.O. and Oguntoke, O., 2020. Public health impacts of improper solid waste management in Africa: a review. Journal of Public Health in Africa, 11(s1).
53. Mehmood, S., Hafeez, S., Shahbaz, M., Abbas, Q., Ahmad, S. and Batool, S.A., 2019. Health and environmental impacts of waste management practices: a review. Journal of Material Cycles and Waste Management, 21(6), pp.1321-1332.
54. Kumar, M., Datta, P., Singh, J. and Mohanty, A., 2019. Solid waste management and its impact on health: A review. Journal of environmental management, 238, pp.471-481.
55. Dikshit, P. and Jain, R., 2019. Health effects of poor waste management in India: A review. International Journal of Environmental Health Research, 29(2), pp.144-158.
56. Gaurav, A. and Singh, A., 2020. Review of health impacts of poor waste management practices in India. Environment and Pollution, 9(3), pp.465-475.
57. Cao, G., Zhang, X., & Gong, S. (2014). The impacts of biomass burning and vehicular emissions on regional climate over the Pearl River Delta, China. Atmospheric Environment, 98, 63-70.
58. Forster, P. M., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D. W., … & Shindell, D. (2007). Changes in atmospheric constituents and in radiative forcing. In Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 129-234). Cambridge University Press.
59. Analysis of Air Pollution due to Vehicle Exhaust Emissions on The Road Networks of Beringin Janggut Area :Achmad Rizki Pratama et al 2019 J. Phys.: Conf. Ser. 1198 082030

60. Harish, M. (2012). A study on air pollution by automobiles in Bangalore city. Management Research and Practice, 4(3), 60-74.
61. Gupta, G. S., & Sharma, S. (2015). Study on impact of open burning on the air quality of Chandigarh city, India. International Journal of Scientific Research, 4(10), 13-15.
62. Kundu, D., & Mandal, T. K. (2017). Impact of vehicular emissions on air quality of a traffic intersection in Kolkata, India. Environmental Monitoring and Assessment, 189(11), 569.
63. Saxena, S., & Guttikunda, S. K. (2010). Preliminary assessment of air pollution and greenhouse gases (GHGs) in Mumbai. Urban Climate News, 37(1), 26-30.
64. Zaman, A. U., & Lehmann, S. (2019). The importance of recycling. Journal of Cleaner Production, 240, 118168
65. Lu et al., 2018
66. Das, N., & Chandran, P. (2011). Microbial degradation of petroleum hydrocarbon contaminants: an overview. Biotechnology research international, 2011, 941810.
67. Lam, S. S., Ujang, Z., & Bakar, S. A. (2012). Waste-to-Energy: A Review of the Status and Benefits in Developed and Developing Countries. Waste Management, 32(10), 1857–1868.
68. Hasan, M. M., Rahman, M. M., & Islam, M. N. (2021). Landfill site suitability analysis using GIS and remote sensing techniques: a case study of Savar municipality, Bangladesh. Journal of Environmental Science and Health, Part A, 56(3), 288-298.
69. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med. 2008 Dec 16;3:17. doi: 10.1186/1751-0473-3-17. PMID: 19087314; PMCID: PMC2633005.