### Codebasics.io

**Primary And Secondary Analysis**

**Primary Analysis (Based on Available data):**

1. List the top 5 and bottom 5 areas with the highest average AQI. (Consider the areas which contain data from the last 6 months: December 2024 to May 2025)
2. List out the top 2 and bottom 2 prominent pollutants for each state of southern India. (Consider data post-COVID: 2022 onwards)
3. Does AQI improve on weekends vs weekdays in Indian metro cities (Delhi, Mumbai, Chennai, Kolkata, Bengaluru, Hyderabad, Ahmedabad, Pune)? (Consider data from the last 1 year)
4. Which months consistently show the worst air quality across Indian states — (Consider the top 10 states with high distinct areas)
5. For the city of Bengaluru, how many days fell under each air quality category (e.g., Good, Moderate, Poor, etc.) between March and May 2025?
6. List the top two most reported diseases or illnesses in each state over the past three years, along with the corresponding average Air Quality Index (AQI) for that period.
7. List the top 5 states with high EV adoption and analyse if their average AQI is significantly better compared to states with lower EV adoption

A Pearson correlation coefficient of 0.11 between the avg\_aqi\_value and the EV\_Count indicates a weak positive linear relationship between average AQI values and electric vehicle (EV) counts, not a negative one. This means that as the EV count increases, the AQI value tends to increase slightly as well, although the association is very weak.

It is a common misconception to expect a negative correlation here because one might intuitively think that more EVs should reduce pollution (and thus AQI). However, the correlation coefficient measures the linear association present in your data. **A positive correlation does not necessarily imply causation or even a meaningful relationship; it just shows the direction of co-movement between variables.**

Since the coefficient is close to zero (0.11), it suggests a very weak or practically no linear relationship, which could also be due to confounding factors, delays in environmental impact, or data characteristics[1](https://www.investopedia.com/ask/answers/032515/what-does-it-mean-if-correlation-coefficient-positive-negative-or-zero.asp)[4](https://www.scribbr.com/statistics/correlation-coefficient/).

**Secondary Analysis (This will require additional data and research)**

1. Which age group is most affected by air pollution-related health outcomes, and how does this vary by city?

Ans:

Children (especially under 5 years):

* **Higher risk of respiratory infections** such as pneumonia, bronchitis, and asthma.
* **Increased vulnerability** due to developing lungs, higher breathing rates, and immature immune systems.
* **Long-term consequences** such as impaired lung development and elevated lifetime risk of respiratory and cardiovascular diseases.
* In low- and middle-income countries, **up to 44% of pneumonia deaths in children under 5** are attributable to household air pollution.
* Urban children, especially those living near high-traffic roads or in industrialized zones, face elevated risks and higher rates of hospitalization due to respiratory diseases.

Older Adults (usually 65 and above):

* **Higher premature mortality** from chronic conditions like heart disease, lung cancer, stroke, and chronic obstructive pulmonary disease. The risk increases substantially for those aged 70 and older.
* Air pollution exacerbates pre-existing chronic illnesses common in this group and increases susceptibility to serious respiratory infections.
* Declining lung function and immune response make the elderly more susceptible not only to the development but also the complication of diseases linked to air pollution.

**City Differences Affect Age Group: Impact within India**

**Indo-Gangetic Plain (e.g., Delhi, Kolkata, Patna, Lucknow):** Extremely high PM2.5 levels

|  |  |  |
| --- | --- | --- |
| Region/City | Most Affected Group | Typical Impact Patterns |
| Delhi | Elderly, Children | Highest adult mortality (elderly); severe child asthma, increased hospital admissions |
| Kolkata | Children | Half of children suffer from respiratory problems, high new-onset asthma |
| Patna, Lucknow, Bihar | Children | High rates of paediatric allergies, infections, and increased infant breathing difficulties |
| Mumbai, Hyderabad | Elderly | Premature adult death from chronic conditions; child risks present, but less severe than north |
| Rural/small towns | Children, Elderly | Major respiratory illnesses in children and the elderly are more common due to both indoor and outdoor pollution |

1. Who are the major competitors in the Indian air purifier market, and what are their key differentiators (e.g., price, filtration stages, smart features)?

Ans:

**Comparative Table of Leading Air Purifier Brands in India**

| **Brand** | **Models (Popular)** | **Price Range (₹)** | **Features & USP** | **Country Origin** |
| --- | --- | --- | --- | --- |
| Coway | AirMega 150, Aim | 10,000–32,000 | True HEPA, Korean design, silent, simple controls | S. Korea/China |
| Dyson | TP10, TP07, BP02 | 32,000–68,900 | HEPA H13, AQI display, airflow tech, app, design | UK/Malaysia/China |
| Philips | AC1215/20, AC1711 | 8,499–25,999 | Multi-stage, real AQI, smart app, allergy relief | Netherlands/China |
| Honeywell | Touch V3, V4, V5 | 4,999–22,999 | HEPA, carbon, UV/plasma, digital & reliable | USA/China/Korea |
| Xiaomi (Mi) | Air 4 Lite, Smart 4 | 10,000–14,000 | Fast PM cleaning, app, smart scenes, budget | China |
| Eureka Forbes | Smart 355, Surround 355 | 9,000–19,000 | 360° intake, WiFi, Indian service | India |
| Sharp | FP-F40E-W, FP-J30M-B | 10,000–18,000 | Ion tech, anti-microbial, low noise | Japan |
| Qubo | Q1000, Q500 | 10,000–20,000 | App/voice, HEPA H13, long filter, ultra-quiet | India |
| Havells | Studio series | 8,000–20,000 | HEPA, carbon, humidifier, design, energy efficient | India |
| Voltas | VAP26TWV/36TMW | 5,000–10,000 | HEPA, carbon, budget-friendly | India |

**AC Air Purifier Affordable Add-Ons:**

| **Product/Brand** | **Filtration Type** | **Compatible ACs** | **Typical Price (₹)** | **Distinguishing Features** |
| --- | --- | --- | --- | --- |
| 3M Filtrete | Electrostatic/HEPA | Split/Window | 300–600 | International brand, affordable, easy use [1](https://indianexpress.com/article/technology/techook/a-look-at-air-conditioners-that-act-as-air-purifiers-7252934/) |
| Nirvana Being Nanotech | Nanotech, multi-layer | Split | 1,495 | Virus removal, lab-tested, urban focus [1](https://indianexpress.com/article/technology/techook/a-look-at-air-conditioners-that-act-as-air-purifiers-7252934/) |
| Finehepa | HEPA H13 | Split/Window | 600–1,100 | High-grade HEPA, simple stick-on [2](https://www.bestreviewsonline.in/air-filters) |
| Prana Air | HEPA, Activated Carbon | Split | 99–799 | Extreme value, 99.7% filtration [3](https://www.pranaair.com/product/ac-filter/) |
| ILoveCleanAir | HEPA | Split | 587+ | Dust/pollen/allergen targeting [4](https://www.flipkart.com/ilovecleanair-ac-filter-transform-your-into-powerful-air-purifier/p/itma671444bb2358) |
| Airth | HEPA + Antiviral | Split/Window/Other | 1,549–2,499 | Plant-based coating, IIT-developed |

**What Is Airth?**

Airth is an Indian company specializing in innovative air purification solutions that *transform existing air conditioning (AC) units into air purifiers*. Instead of selling standalone purifiers, Airth offers add-on modules—primarily HEPA filter-based—that are easy to install on various AC types, including split, window, and cassette ACs. Developed through research from top institutes such as IIT Kanpur and IISc Bangalore, Airth aims to make clean air affordable and accessible in Indian homes and offices[1](https://airth.in/)[2](https://airth.in/products/split-ac-air-purifier)[3](https://www.1mg.com/otc/airth-air-purifier-for-ac-summer-version-aqi-200-otc1041336).

**Product Lineup**

| **Product Type** | **Compatible With** | **Indicative Price Range** | **Key Features** |
| --- | --- | --- | --- |
| Split AC Air Purifier | All Split ACs | ₹2,499[1](https://airth.in/)[4](https://price-history.in/product/airth-air-purifier-ac-hepa-ac-FxRRxjce) | HEPA filter, 99% dust/allergen/bacteria removal |
| Window AC Air Purifier | Window ACs | ₹1,549[1](https://airth.in/) | HEPA filter, compact design |
| Cassette AC Air Purifier | Cassette ACs | ₹1,999[1](https://airth.in/) | Custom fit, HEPA filter |
| Activated Carbon version | Split/Window ACs | ~₹2,499–₹4,000[4](https://price-history.in/product/airth-air-purifier-ac-hepa-ac-FxRRxjce)[5](https://airth.in/products/air-purifier-for-ac-with-activated-carbon) | Enhanced with VOC/gas removal |
| Vsure Room Purifier | Standalone unit | ₹5,500–₹49,999[6](https://www.indiamart.com/proddetail/airth-vsure-air-purifier-26018187048.html) | HEPA filter, larger area coverage |

1. What is the relationship between a city’s population size and its average AQI — do larger cities always suffer from worse air quality? (Consider 2024 population and AQI data for this)

Ans: There is almost no correlation between a city’s population size and its average AQI, which is backed by the Pearson’s correlation coefficient factor that comes out to be 0.09.

Demand for air purifiers in India is highly concentrated in urban centres with severe air pollution, especially in North India**.**

**Cities with the worst air quality, rising health awareness, and high disposable incomes see the greatest uptake. The spike in demand is most pronounced during winter and high-smog seasons**.

| **City** | **Demand Level** | **Key Drivers** | **Seasonality** |
| --- | --- | --- | --- |
| Delhi NCR | Very High | Severe air pollution, health events, and public awareness | Peaks Oct-Dec |
| Mumbai | High/Rising | Recurring smog, rising AQI awareness, and health focus | Smog/winter peaks |
| Bengaluru | High | High incomes, pollution, and tech-savvy consumers | Year-round |
| Chennai | Moderate/High | Seasonal, urban growth, health trends | Smog/winter |
| Hyderabad | Moderate/High | Rising pollution, health-conscious residents | Winter peak |
| Kolkata | Moderate/High | Seasonal smog, rising urban adoption | Smog/winter |
| Pune | High | Urban migration, the proactive health segment | Mild seasonality |
| Ahmedabad | Moderate/High | Industrial pollution, increased awareness | Smog/winter |
| Tier-II Cities | Growing | Online sales surge, pollution spikes, media alerts | Episodic |

1. How aware are Indian citizens of what AQI (Air Quality Index) means, and do they understand its health implications?

Ans:

| **Finding** | **Statistic/Impact** | **Population Most Affected** | **Study Regions** |
| --- | --- | --- | --- |
| Every 10 AQI increase | 1–5% more asthma ER visits | Children, elderly | US, Europe, Asia |
| High PM2.5/O₃/NO₂ levels | More symptoms, attacks, and hospital use | Urban, low-SES | Global |
| Long-term high AQI | Higher asthma prevalence | All ages, children | Global |
| AQI spikes | Short-term surges in hospital visits | At-risk, urban populations | Global |

| **Report Attribute** | **Key Statistics** |
| --- | --- |
| **Base Year** | 2024 |
| **Forecast Years** | 2025-2033 |
| **Historical Years** | 2019-2024 |
| **Market Size in 2024** | USD 507.5 Million |
| **Market Forecast in 2033** | USD 1,813.5 Million |
| **Market Growth Rate (2025-2033)** | 14.06% |

<https://www.imarcgroup.com/india-air-purifier-market>

An **air purifier** is a device designed to improve indoor air quality by removing contaminants such as dust, [pollen](https://www.imarcgroup.com/pollen-allergy-market), pet dander, smoke, and other airborne particles. It utilizes various technologies, including filters, ionizers, and UV light, to capture or neutralize pollutants, preventing them from circulating in the air. The primary component is the filter, which traps particles as air passes through.

**High-efficiency particulate air (HEPA)** filters are commonly used, capable of capturing tiny particles. Some models also incorporate activated carbon filters to absorb odours and chemicals. Ionizers release negatively charged ions to attract and neutralize airborne particles, while UV light can destroy certain microorganisms. Air purifiers are beneficial for individuals with allergies, asthma, or respiratory issues, as well as for general wellness by creating a cleaner and healthier indoor environment. Regular maintenance, including filter replacement, is essential to ensure optimal performance.

**Filter Technology Insights:**

* High Efficiency Particulate Air (HEPA)
* Activated Carbon
* Ionic Filters
* Others

1. Which pollution control policies introduced by the Indian government in the past 5 years have had the most measurable impact on improving air quality, and how have these impacts varied across regions or cities?

Ans:

**Summary Table: Key National Policies (2020–2025)**

| **Policy/Initiative** | **Description & Focus** | **Timeline** |
| --- | --- | --- |
| NCAP | Reduce PM, city action | 2019–2026 |
| CAQM | NCR-wide emission mgmt | 2020 onward |
| GRAP | Emergency response protocol | Updated yearly |
| BS-VI Standards | Vehicle emission norms | 2020 launch |
| FAME II | Electric vehicle boost | 2020–2025 |
| PMUY | LPG for clean cooking | Ongoing |
| NAMP Expansion | Real-time AQI, more cities | 2020–2025 |
| Industrial Emissions | Tighter limits, plant retrofits | 2021–2025 |
| Crop Residue Mgmt | Curb stubble burning | 2021–2025 |
| SWM Rules (Update) | Waste burning bans, segregation | 2021–2025 |

**Most Impactful Recent Indian Policies on Air Quality Improvement (2020–2025)**

| **Policy** | **Measurable Impact Summary** |
| --- | --- |
| National Clean Air Programme (NCAP) | 51 cities > 20% PM10 reduction; 21 cities >40% PM10 reduction; improved public health |
| BS-VI Emission Norms | Up to 80% reduction in PM, 70% in NOx for new vehicles, directly lowering vehicular pollution |
| GRAP + CAQM | Notable short-term AQI and PM10 reduction in NCR during GRAP enforcement |
| FAME II | Early-stage reduction in fleet emissions, EV adoption trendline improvement |

**Extra Details:**

1. **Answer Critical Questions:** 
   * *Priority Cities*: Which Tier 1/2 cities show irreversible AQI degradation?

**Summary Table: Cities with Most Severe and Potentially Irreversible AQI Degradation**

| **Tier** | **City** | **AQI Trend** | **Primary Causes** | **Evidence of Irreversibility** |
| --- | --- | --- | --- | --- |
| 1 | Delhi NCR | Severe & persistent | Vehicles, industry, stubble burning | Permanent lung damage, chronic public health crisis[1](https://en.wikipedia.org/wiki/Air_pollution_in_Delhi)[2](https://oizom.com/top-10-most-polluted-cities-in-the-world/)[3](https://www.ijcrt.org/papers/IJCRT2410096.pdf) |
| 1 | Kolkata | Unhealthy & worsening | Vehicles, industry, waste burning | Chronic health impacts |
| 1 | Mumbai | Unhealthy in winter/rising | Vehicles, industry, construction | Increasing respiratory burden |
| 1 | Bengaluru | Declining, “Unhealthy” spikes | Urbanization, traffic | Persistent unhealthy status |
| 2 | Ghaziabad/Noida | Severe, continual spikes | Industry, traffic, Delhi proximity | Top polluted globally, no sustained relief[2](https://oizom.com/top-10-most-polluted-cities-in-the-world/)[4](https://www.aqi.in/in/dashboard/india) |
| 2 | Bhiwadi | Severe, industrial & year-round | Heavy industries, vehicles | Worker health, child lung impact |
| 2 | Patna | Unhealthy, frequent spikes | Open burning, construction | No improvement trend |
| 2 | Faridabad | Persistent unhealthy levels | Industry, road dust, traffic | Chronic health impacts |
| 2 | Jaipur | Growing & persistent decline | Construction, vehicles | Warnings about irreversible harm |
| 2 | Indore, Surat | At risk, declining | Urbanization, industry | Policy warnings issued |
| 2 | Davanagere, Kalaburagi, Hubballi-Dharwad | Projected sharp rise | Industry, vehicles | 40% rise warned by 2030 |

* + *Health Burden*: How do AQI spikes correlate with paediatric asthma admissions?

**Key Findings**

* **Strong Positive Correlation:** Spikes in the Air Quality Index (AQI)—especially elevated levels of PM2.5, PM10, NO₂, ozone, and CO—are consistently associated with sharp increases in paediatric asthma emergency room visits and hospital admissions[1](https://pmc.ncbi.nlm.nih.gov/articles/PMC12251416/)[2](https://www.atsjournals.org/doi/10.1513/AnnalsATS.202105-539OC)[3](https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2023.1105140/full).
* **Timing:** The rise in asthma admissions often occurs on the same day as pollution spikes or is delayed by 2–4 days, reflecting the typical “lag effect” from air exposure to respiratory symptoms[2](https://www.atsjournals.org/doi/10.1513/AnnalsATS.202105-539OC)[4](https://www.tandfonline.com/doi/full/10.1080/02770903.2021.2008429)[5](https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.865798/full).
  + *Behaviour Shifts*: Do pollution emergencies increase purifier searches/purchases?

**Trends in Searches and Consumer Interest**

**Spike in Online Searches:** During pollution emergencies (e.g., post-Diwali smog in Delhi, winter haze in North India), Google Trends and e-commerce data consistently show a dramatic rise in searches for terms **like “air purifier,” “best air purifier,” and “HEPA filter.”**

Search interest can increase by 2–5x within days of a **major smog episode**.

**Seasonal Peaks:** The highest spikes in air purifier searches typically align with key pollution periods: **late October through January in North Indian cities**, and **during local pollution events in Tier 1 and Tier 2 markets**.

**Media/Alert Influence:** AQI alerts from government agencies and prominent media reports about hazardous air have often led to rapid increases in interest.

**Effects on Air Purifier Purchases**

**Short-term Demand Surges**: Retailers and e-commerce platforms frequently report sales surges **of 40–200% within a week of an AQI emergency.** **In Delhi NCR**, purifier stockouts are common after major smog episodes.

**Expansion Beyond Metros:** Tier 2 cities—such as **Lucknow, Kanpur, Surat, and Bhubaneswar**—increasingly show sharp sales bumps after local AQI crises.

**Repeat Behaviour:** This pattern repeats each year, often overwhelming supply, especially with increased health awareness post-pandemic.

| **Pollution Emergency** | **Search Interest** | **Purchase Spike** |
| --- | --- | --- |
| Diwali Smog (Delhi) | 3–5x jump | 2–3x sales increase, frequent stockouts |
| Post-monsoon/winter (NCR) | 2–4x spike | 50–150% sales boost |
| AQI Alerts (Tier 2 cities) | Notable rise | 40–100% bump observed by retailers |

* + *Feature Gap*: What do existing products lack (e.g., smart AQI syncing, compact designs)?

| **Feature Area** | **Current Gap** |
| --- | --- |
| Smart AQI Syncing | Limited real-time outdoor AQI link, lack of predictive automation |
| Design & Portability | Few truly compact or modular options for urban/small spaces |
| Smart Home Integration | Siloed apps, limited interoperability with the broader smart home |
| User Education & Insights | Minimal actionable info or personalized health alerts |
| Maintenance | No real-time filter wear detection, limited eco-friendly options |
| Customization | Poor adaptation to different room sizes/layouts |
| Noise & Power Efficiency | Insufficient low-noise, energy-transparent models |

1. **Deliverables:**

**Market Prioritization Dashboard** with:

* City risk scores (AQI severity × population density × income)
* Health cost impact projections
* Competitor feature gap matrix

**Product Requirements Document** specifying:

* Must-have features (e.g., PM2.5/VOC sensors)
* Tiered pricing models for target segments

1. **Innovate:** 
   * + - * Integrate external data (e.g., Google Trends, crop-burning satellite imagery)
         * Video must demonstrate dashboard functionality + city-specific entry simulations

Data cleaning steps: 15/07/2025, 11 – 11.30 pm

**MYSQL**

1. File Import large CSV with ~ 1 M rows.

Method 1 (lengthy) – Import wizard

Method 2 (fast) – Load data in local file (create table with all columns except the last column, i.e., note, because it is mostly empty)

1. Change the datatypes of dates from text type to date type, convert empty spaces into Nulls, then modify datatypes

Big Data import:

create database airpurifier;

use airpurifier;

create table vahan (

year int

, month text

, state text

, rto text

, vehicle\_class text

, fuel text

, value int

, unit text

);

LOAD DATA LOCAL INFILE 'D:/CN/Hackathons/Code Basics/vahan.csv'

INTO TABLE vahan

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\r\n'

IGNORE 1 LINES;

Primary Analysis: **Sat -** 19/07/2025 to **Sun -** 21/07/2025

Secondary Analysis: 22/07/2025

Extra Details: 23/07/2025

Other Data Sourcing & Data modelling in Power BI: 24/07/2025 – 25/07/2025

Dashboarding in Power BI:

26/07/2025 – Data Modelling

27/07/2025 – decide KPIs

28/07/2025 – Data cleaning further, adding a few more tables to the data model

**EXCEL**

1. Two states from the “aqi” table were having a common area named “Aurangabad,” which I edited to differentiate according to their native states 1. Aurangabad (MH) and the other 2. Aurangabad (BR), where MH = Maharashtra and BR = Bihar.
2. The states and UTs were misspelled in the IDSP table, which I corrected using the switch function

= SWITCH(A2,

"Arunachal", "Arunachal Pradesh",

"Madhya", "Madhya Pradesh",

"Uttar", "Uttar Pradesh",

"Dadra and Nagar Haveli", "Dadra and Nagar Haveli and Daman and Diu",

A2)

1. Transformed idsp table by correcting disease\_illness\_name column rows by correcting misspelled names, using the Xlookup formula:  
     
   = XLOOKUP(G2, Sheet1!A$2:A$123, Sheet1!B$2:B$123, G2)

Where

1. idsp!G2:G6475 contains the column disease\_illness\_name with 6475 lakh entries.
2. Sheet1!A2:A123 contains the original (uncorrected) unique disease names.
3. Sheet1!B2:B123 contains the corrected names.
4. Created a subtable of idsp that only contains the respiratory diseases.

**POWER BI**

1. Create a table named “Measure KPIs” entirely dedicated to all the measures.
2. Import aqi csv to powerbi – transform data -> delete note column -> add a column called “pollutant\_count” adjacent to the column “prominent\_pollutants”.

DAX:  
pollutant\_count =

IF(

    ISBLANK(aqi[prominent\_pollutants]),

    0,

    LEN(TRIM(aqi[prominent\_pollutants]))

    - LEN(SUBSTITUTE(TRIM(aqi[prominent\_pollutants]), ",", ""))

    + 1

)

1. Import idsp csv – transform -> delete note column -> found 3 error rows

Created a copy of idsp csv to an Excel file – transform -> load -> delete note column (No Error)

1. The source tables aqi, idsp, vahan, and population\_projection did not have a unique ID **(the** **common unique identifier**) in any of them, which directed me to create a **Many-to-Many** relationship cardinality.

To avoid ambiguity while filtering, I created five bridge tables: **Year, Month, Week of the Year, State or UT, and Area or City**. This would allow me to create one-to-many relationships with the other dimension tables.

1. Top KPIs –
   * 1. The most common pollutant
     2. The most polluted city with its AQI
     3. The most common feature in the air purifiers
     4. City with the highest disposable income
     5. City with the highest liveability score
     6. City with the maximum EV adoption
     7. City Risk Score = AQI severity × population density × income

Visuals:

1. Multi-row card:
2. Ribbon chart:
3. Card:
4. 100% stacked column (Instead of Gauge):