Problem Statement

Conduct Market Fit Research for Air Purifier Development Using AQI Analytics

Domain: Consumer Appliances

Function: Market Research Analytics

"**AirPure Innovations**" is a startup born out of the air quality crisis in India, with 14 cities ranking among the world's top 20 most polluted urban centers. The company is in the early stages of product development and is unsure whether there is a strong, sustained demand for its air purifier product. Before committing to production and R&D, they need to answer critical questions:

1. What pollutants or particles should their air purifier target?
2. What are the most essential features that should be incorporated into the air purifier?
3. Which cities have the highest demand for air purifiers, and what is the market size in these regions?
4. How can R&D be aligned with localized pollution patterns?

The urgency is highlighted by real-world examples: biotech entrepreneur Bryan Johnson walked out of a podcast due to poor air quality, and Delhi's Taj Hotels now display AQI readings, emphasizing the growing awareness of air quality in daily life and business decisions.

COO **Tony Sharma** believes success lies in analyzing three key dimensions of the market. Tony has reached out to Peter Pandey, a Data Analyst, to assist in gathering insights for strategic decisions. The three dimensions to be analyzed are:

1. **Severity Mapping**: Identify cities experiencing persistent or worsening AQI (Air Quality Index) levels.
2. **Health Impact Correlation**: Quantify the health burden due to pollution and its impact on consumers' well-being.
3. **Demand Triggers**: Examine the relationship between pollution spikes and shifts in consumer behavior related to air purifier demand.

While searching for reliable, real-time AQI data, the team discovered the "Dataful" platform, which provides datasets essential for answering these questions.

Task

Imagine yourself as **Peter Pandey** and perform the following tasks:

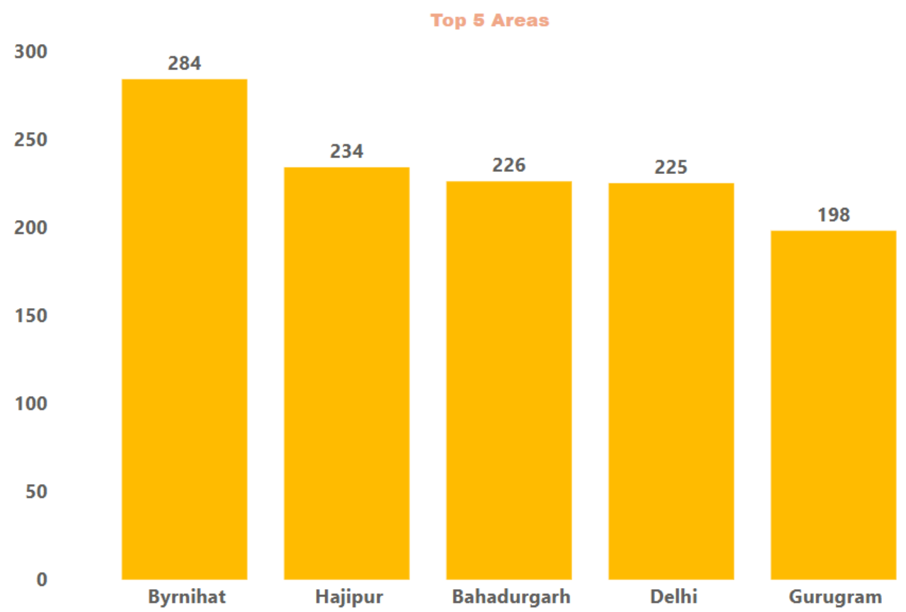
- Check 'primary_and_secondary_questions.pdf'. You can use any tool of your choice (Python, SQL, Power BI, Tableau, Excel, PowerPoint) to analyse and answer these questions. More relevant instructions are provided in this document.
- Design a dashboard with your metrics and analysis. The end users of this dashboard are top-level management and the product strategy team - hence, the dashboard should be self-explanatory and easy to understand.
- Present your insights to Tony Sharma & team. Be creative and concise with your presentation.
- Use your dashboard in the presentation along with the deck.
- Use additional data based on your own research to support your recommendations.

Primary Analysis

1. List the top 5 and bottom 5 areas with highest average AQI. (Consider areas which contains data from last 6 months: December 2024 to May 2025)

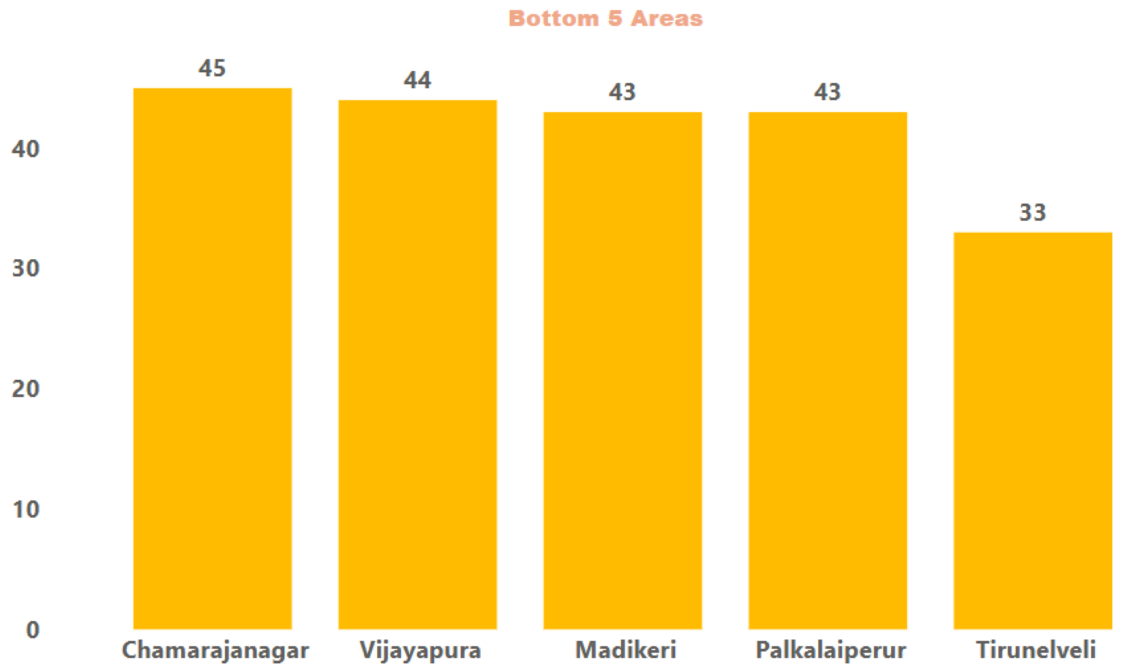
Top 5 Areas with highest average AQI :

```sql
SELECT
    area, ROUND(AVG(aqi_value), 0) AS avg_aqi
FROM
    aqi
WHERE
    date BETWEEN '2024-12-01' AND '2025-04-30'
GROUP BY area
ORDER BY avg_aqi DESC
LIMIT 5;
```



Bottom 5 Areas with highest average AQI :

```sql
SELECT
    area, ROUND(AVG(aqi_value), 0) AS avg_aqi
FROM
    aqi
WHERE
    date BETWEEN '2024-12-01' AND '2025-04-30'
GROUP BY area
ORDER BY avg_aqi
LIMIT 5;
```



2. List out top 2 and bottom 2 prominent pollutants for each state of southern India. (Consider data post covid: 2022 onwards)

Top 2 prominent pollutants :

```
WITH top_prominent_pollutants AS (
    SELECT
        state,
        prominent_pollutants,
        COUNT(prominent_pollutants) AS p_count,
        DENSE_RANK() OVER (
            PARTITION BY state
            ORDER BY COUNT(*) DESC
        ) AS ranking
    FROM
        aqi
    WHERE
        YEAR(date) >= 2022
        AND state IN (
            "Andhra Pradesh",
            "Kerala",
            "Karnataka",
            "Tamil Nadu",
            "Telangana"
        )
    GROUP BY
        state,
        prominent_pollutants
)

SELECT
    *
FROM
    top_prominent_pollutants
WHERE
    ranking <= 2;
```

Top 2 Pollutants

State	Pollutants
Andhra Pradesh	PM10, PM2.5
Karnataka	PM10, CO
Kerala	PM10, PM2.5
Tamil Nadu	PM10, PM2.5
Telangana	PM10, PM2.5

Bottom 2 prominent pollutants :

```
WITH bottom_prominent_pollutants AS (
    SELECT
        state,
        prominent_pollutants,
        COUNT(prominent_pollutants) AS p_count,
        DENSE_RANK() OVER (
            PARTITION BY state
            ORDER BY COUNT(*) ASC
        ) AS ranking
    FROM
        aqi
    WHERE
        YEAR(date) >= 2022
        AND state IN (
            "Andhra Pradesh",
            "Kerala",
            "Karnataka",
            "Tamil Nadu",
            "Telangana"
        )
    GROUP BY
        state,
        prominent_pollutants
)

SELECT
    *
FROM
    bottom_prominent_pollutants
WHERE
    ranking <= 2;
```

Bottom 2 Pollutants

State	Pollutants
Andhra Pradesh	SO_2, NO_2
Karnataka	SO_3, NH_3
Kerala	SO_2, NH_3
Tamil Nadu	NH_3, NO_2
Telangana	NO_2, O_3

3. Does AQI improve on weekends vs weekdays in Indian metro cities (Delhi, Mumbai, Chennai, Kolkata, Bengaluru, Hyderabad, Ahmedabad, Pune)? (Consider data from last 1 year)

Avg. AQI on Weekday & Weekend (by Indian Metro Cities):

```sql
with new_date as(select max(date) as max_date from aqi)
select area,
case
    when dayofweek(date) in (1,7) then "Weekend"
    else "Weekday"
end as day_type,
round(avg(aqi_value),0) as avg_aqi from aqi,new_date
where area in ("Delhi", "Mumbai", "Chennai", "Kolkata",
"Bengaluru", "Hyderabad", "Ahmedabad", "Pune") and date
>= new_date.max_date - interval 1 year
group by area,day_type
order by area,day_type;
```

Avg_AQI on Weekday & Weekend

area	Weekday	Weekend
Ahmedabad	109	111
Bengaluru	71	73
Chennai	69	67
Delhi	192	184
Hyderabad	79	79
Kolkata	82	83
Mumbai	89	88
Pune	103	100

Avg. AQI on Weekday & Weekend:

```
with new_date as(select max(date) as max_date from aqi)
select
case
    when dayofweek(date) in (1,7) then "Weekend"
    else "Weekday"
end as day_type,
round(avg(aqi_value),0) as avg_aqi from aqi,new_date
where date >= new_date.max_date - interval 1 year
group by day_type
order by day_type;
```

Day_Type	Avg_AQI
Weekday	101
Weekend	99

4. Which months consistently show the worst air quality across Indian states — (Consider top 10 states with high distinct areas)

Top 10 states with high distinct areas:

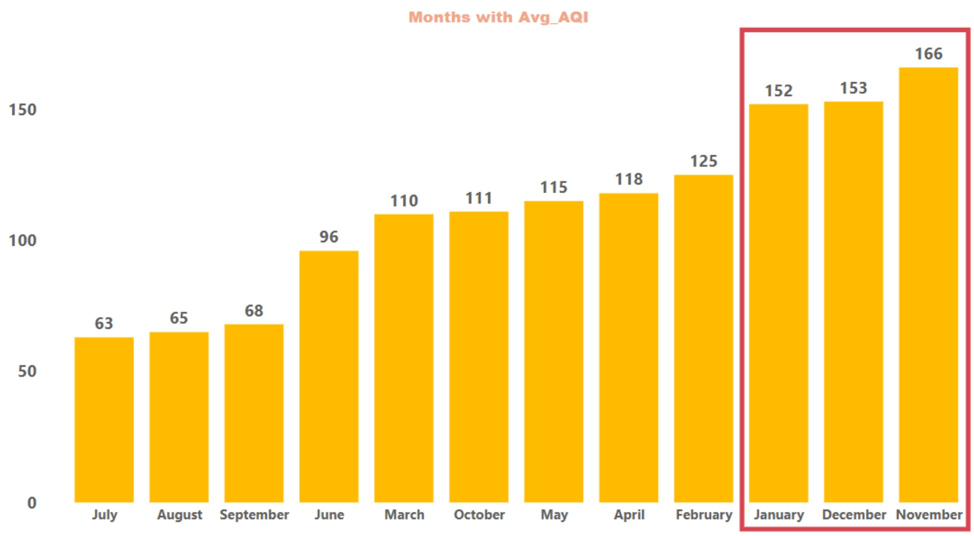
```
SELECT distinct state,count(distinct area) as
total_monitoring_stations
from aqi
group by state
order by total_monitoring_stations desc
limit 10;
```

States & Monitoring Stations

state	Monitoring_Stations
Rajasthan	34
Maharashtra	31
Karnataka	27
Tamil Nadu	26
Bihar	25
Haryana	25
Uttar Pradesh	20
Odisha	16
Madhya Pradesh	15
Andhra Pradesh	9

Months consistently show the worst air quality:

```sql
WITH high_state AS (
    SELECT
        state,
        COUNT(DISTINCT area) AS c
    FROM
        aqi
    GROUP BY
        state
    ORDER BY
        c DESC
    LIMIT 10
)

SELECT
    DATE_FORMAT(a.date, '%M') AS month_name,
    ROUND(AVG(a.aqi_value), 0) AS avg_aqi_value
FROM
    aqi a
JOIN
    high_state h ON a.state = h.state
WHERE
    a.state IN (SELECT state FROM high_state)
GROUP BY
    month_name
ORDER BY
    avg_aqi_value DESC;
```



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?

```sql
SELECT
    area,air_quality_status, COUNT(DISTINCT date)
AS "no. of days"
FROM
    aqi
WHERE
    area = "Bengaluru"
        AND date >= "2025-03-01"
GROUP BY air_quality_status;
```

air_quality_status	no. of days
Moderate	13
Satisfactory	48

6. List the top two most reported disease illnesses in each state over the past three years, along with the corresponding average Air Quality Index (AQI) for that period.

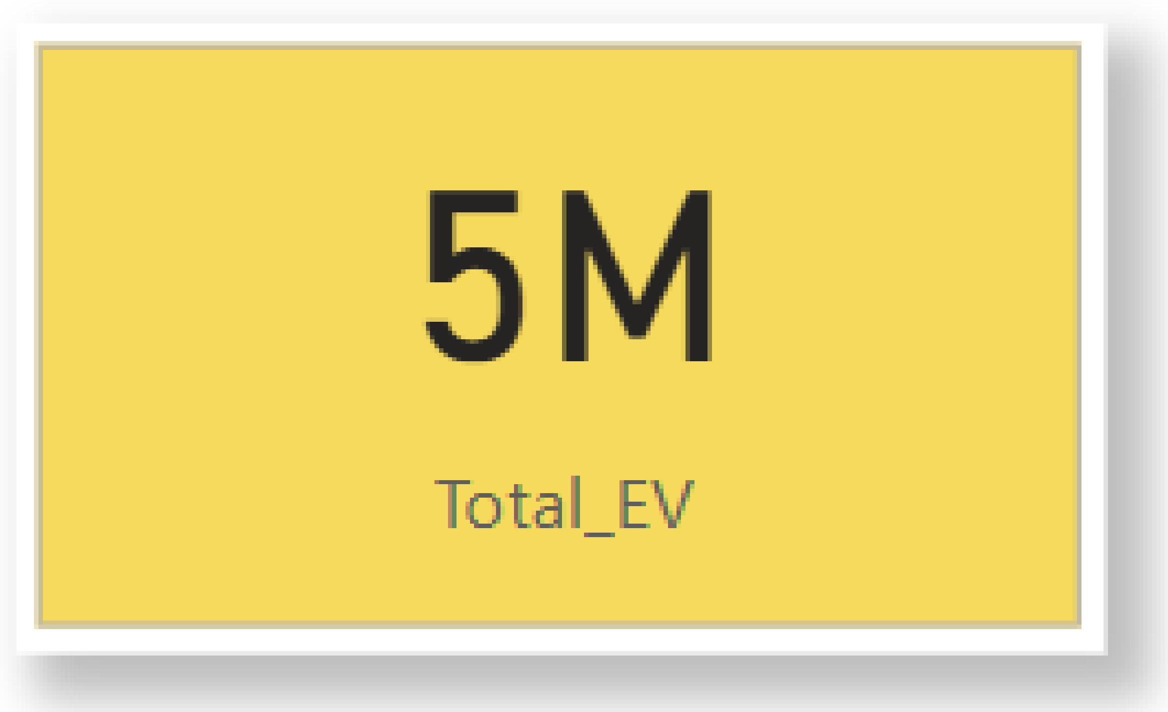
```sql
WITH disease_data AS (
    SELECT
        disease_illness_name,
        state,
        reporting_date,
        SUM(cases) AS daily_cases
    FROM
        idsp
    WHERE
        reporting_date >= DATE_SUB('2025-04-30', INTERVAL 3 YEAR)
    GROUP BY
        disease_illness_name, state, reporting_date
),

aqi_state_avg AS (
    SELECT
        state,
        ROUND(AVG(aqi_value), 0) AS avg_aqi
    FROM
        aqi
    WHERE
        date >= DATE_SUB('2025-04-30', INTERVAL 3 YEAR)
    GROUP BY
        state
),

disease_totals AS (
    SELECT
        disease_illness_name,
        state,
        SUM(daily_cases) AS total_cases
    FROM
        disease_data
    GROUP BY
        disease_illness_name, state
),

ranked_diseases AS (
    SELECT *,
           RANK() OVER (PARTITION BY state ORDER BY total_cases DESC) AS disease_rank
    FROM disease_totals
),

pivoted AS (
    SELECT
        state,
        MAX(CASE WHEN disease_rank = 1 THEN disease_illness_name END) AS top_disease_1,
        MAX(CASE WHEN disease_rank = 2 THEN disease_illness_name END) AS top_disease_2
    FROM
        ranked_diseases
    WHERE
        disease_rank <= 2
    GROUP BY
        state
)

SELECT
    p.state,
    p.top_disease_1,
    p.top_disease_2,
    a.avg_aqi
FROM
    pivoted p
JOIN
    aqi_state_avg a ON p.state = a.state
ORDER BY
    state;
```

state	top_disease_1	top_disease_2	avg_aqi
Andaman and Nicobar Islands	Acute Diarrheal Disease	Fever with Rash	58
Andhra Pradesh	Acute Diarrheal Disease	Cholera	77
Arunachal Pradesh	Acute Diarrheal Disease	Chickenpox	54
Assam	Acute Diarrheal Disease	Food Poisoning	111
Bihar	Acute Diarrheal Disease	Dengue	153
Chandigarh	Cholera	NULL	133
Chhattisgarh	Acute Diarrheal Disease	Cholera	78
Delhi	Dengue	Measles	188
Gujarat	Acute Encephalitic Syndrome (Chandipura Virus)	Acute Diarrheal Disease	110
Haryana	Cholera	Acute Diarrheal Disease	140
Himachal Pradesh	Acute Diarrheal Disease	Hepatitis A	160
Jammu and Kashmir	Dengue	Hepatitis A	71
Jharkhand	Malaria	Acute Diarrheal Disease	164
Karnataka	Acute Diarrheal Disease	Cholera	63
Kerala	Food Poisoning	Acute Diarrheal Disease	67
Madhya Pradesh	Acute Diarrheal Disease	Dengue	107
Maharashtra	Acute Diarrheal Disease	Food Poisoning	103
Manipur	Dengue	Food Poisoning	103
Meghalaya	Measles	Acute Diarrheal Disease	62
Mizoram	Food Poisoning	Scrub Typhus	47
Nagaland	Dengue	Acute Diarrheal Disease	80
Odisha	Acute Diarrheal Disease	Food Poisoning	124
Puducherry	Acute Diarrheal Disease	Dengue	57
Punjab	Acute Diarrheal Disease	Cholera	117
Rajasthan	Acute Diarrheal Disease	Dengue	128
Sikkim	Jaundice	Typhoid	54
Tamil Nadu	Acute Diarrheal Disease	Mumps	68
Telangana	Acute Diarrheal Disease	Food Poisoning	79
Tripura	Acute Diarrheal Disease	Dengue	125
Uttar Pradesh	Acute Diarrheal Disease	Food Poisoning	119
Uttarakhand	Dengue	Acute Diarrheal Disease	88
West Bengal	Acute Diarrheal Disease	Food Poisoning	109

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

Total EVs:

```
select sum(value) as Total_EV from vahan
where fuel in (
"ELECTRIC(BOV)",
  "PURE EV",
  "PLUG-IN HYBRID EV",
  "STRONG HYBRID EV",
  "SOLAR",
  "FUEL CELL HYDROGEN");
```



Top 5 States with high EV Adoption:

```sql
WITH total_ev AS (
  SELECT
    state,
    SUM(value) AS total_EVs,
    RANK() OVER (ORDER BY SUM(value) DESC) AS ev_rank
  FROM vahan
  WHERE fuel IN (
    "ELECTRIC(BOV)",
    "PURE EV",
    "PLUG-IN HYBRID EV",
    "STRONG HYBRID EV",
    "SOLAR",
    "FUEL CELL HYDROGEN"
  )
  GROUP BY state
),
aqi_avg AS (
  SELECT state, ROUND(AVG(aqi_value), 0) AS avg_aqi
  FROM aqi
  GROUP BY state
)
SELECT
  t.state,
  t.total_EVs,
  a.avg_aqi
FROM total_ev t
JOIN aqi_avg a ON t.state = a.state
ORDER BY total_EVs desc
limit 5;
```

States with High EV

state	total_EVs	avg_aqi
Uttar Pradesh	921471	121
Maharashtra	650823	103
Karnataka	480191	63
Tamil Nadu	329634	68
Rajasthan	305605	128

States with lower EV Adoption:

```sql
WITH total_ev AS (
  SELECT
    state,
    SUM(value) AS total_EVs,
    RANK() OVER (ORDER BY SUM(value) DESC) AS ev_rank
  FROM vahan
  WHERE fuel IN (
    "ELECTRIC(BOV)",
    "PURE EV",
    "PLUG-IN HYBRID EV",
    "STRONG HYBRID EV",
    "SOLAR",
    "FUEL CELL HYDROGEN"
  )
  GROUP BY state
),
aqi_avg AS (
  SELECT state, ROUND(AVG(aqi_value), 0) AS avg_aqi
  FROM aqi
  GROUP BY state
)
SELECT
  t.state,
  t.total_EVs,
  a.avg_aqi
FROM total_ev t
JOIN aqi_avg a ON t.state = a.state
where t.state in
("Jharkhand","Tripura","Chandigarh","Himachal
Pradesh","Manipur")
ORDER BY total_EVs desc
limit 5;
```

States with Low EV

state	total_EVs	avg_aqi
Chandigarh	17888	134
Himachal Pradesh	4963	160
Jharkhand	67695	164
Manipur	1394	102
Tripura	22631	123

Secondary Analysis

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

-- Children, especially newborns, are most affected by air pollution. (1-5)

-- Indoor pollution (dirty fuels, no kitchen) increases death risk.

-- In polluted cities (like Delhi & Kolkata), child mortality risk is ~2x higher.

-- Adults are also affected, but less than children.

-- Southern & Northeast states show lower risk due to cleaner air and fuels.

(Why Southern & Notheast states are low polluted.)→ answer this in vdo

→ Source: "Air Pollution and Mortality in India," GeoHealth, 2024 (NFHS-5 based)

→ link: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GH000968

In 2024, approximately 464 children under the age of 5 died every day in India due to air pollution. This amounts to an estimated annual total of around 169,000 child deaths attributed directly to air pollution-related causes.

A trusted source for this data is the "State of Global Air 2024" report, released by the Health Effects Institute (HEI) in partnership with UNICEF, and widely reported in major Indian news outlets123.

1. https://economictimes.com/news/india/air-pollution-every-day-464-children-in-india-die-report/articleshow/111133693.cms

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

Brand	**Most Popular Model**	**Price (₹)**	**Filtration Stages**	**Smart Features**	**Coverage Area (sq.ft)**
Philips	1000i Series Smart (AC1715/60)	~14,995	3 (HEPA, Active Carbon, Pre-filter)	Voice control (Alexa, Google), App, Sleep/Auto mode	Up to 387
Daikin	MC55XVM6 Active Plasma	~16,990	3 (Electrostatic HEPA, Plasma Ion, Deodorizing)	Auto fan, Econo, Turbo, Child lock, Air quality sensor	441
Honeywell	HAC35M1101W	~18,900	3 (Pre-filter, HEPA, HiSiv/Carbon)	Touch panel, Auto mode, Sleep mode, Filter indicator	450
Blue Star	BS-AP250RAP	~13,000	3+ (HEPA, Carbon, SensAir, Pre-Filter)	PM2.5 LED, SensAir tech, Auto, IAQ monitoring	250–490 (varies)
Blueair	Blue Pure 211	~19,635	3 (Pre-filter, HEPASilent, Carbon)	Filter indicator, Silent mode, Simple controls	540
Sharp	PureFit FX-S120	~51,999	3 (Plasmacluster Ion, HEPA, Carbon)	App control, AIoT, Multi-sensor, Auto, Quiet mode	Unlisted (large rooms)
Panasonic	F-PXM35ASD	~15,990	3 (HEPA, Activated Carbon, Pre-filter)	Auto mode, Sleep, Odor sensors, Child lock	283
Xiaomi	Smart Air Purifier 4 Lite	~12,000	3 (HEPA, Activated Carbon, Pre-filter)	Voice, App, LED display, Auto, Filter monitor	269–463
Dyson	Purifier Cool TP10	~45,000	2–3 (HEPA H13, Carbon)	App/voice, Auto mode, Air quality sensors, Oscillation, Cooling	400
Kent	Aura	~12,500	3 (HEPA, Carbon, UV)	Timer, Child lock, Air quality indicator	290 (typical)

Q3: 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)

Rank	City	2024	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	Byrnihat, India	128.2	174.4	183.7	153.3	140	105.7	88.4	90	86	68.6	87.4	146	193.7
2	Delhi, India	108.3	219.3	110.3	76.8	64.6	84.4	52.6	39.6	26.6	42.8	120.1	276.5	185.9
3	Mullanpur, India	102.3	123.3	111.2	79.6	47.2	67	53.1	72.3	42.5	62.2	142.7	302.5	146
4	Faridabad, India	101.2	211.3	116	107.4	130.9	188.9	100.3	54.2	29.7	36.8	57.8	119.8	74.3
5	Loni, India	91.7	135.7	81.2	60.1	56.9	68.7	63.4	54.7	32.9	58.5	134.5	235.7	120.1
6	New Delhi, India	91.6	190.4	91	66.4	52.8	75.4	49.9	34.6	24.5	37.8	99.6	224	152.1
7	Gurugram, India	87.4	138.8	104.8	86.2	96.6	86.2	65.6	49.6	39.4	43.2	79.9	146.2	113.8
8	Ganganagar, India	86.6	159.4	104.8	64.6	65.7	102.1	78.6	53.9	55.8	69.9	52.5	127.3	93.6
9	Greater Noida, India	83.5	168.9	108.5	72.8	68.6	92.1	59.2	37.6	26	42	87.5	135.7	105.1
10	Bhiwadi, India	83.1	128.2	91.7	74.4	78.7	103.4	77.7	49.7	33.8	40.2	88.3	142.2	85.7

- **Larger cities often have high AQI due to traffic, industry, and urbanization — but city size is not the only factor.**
- **Larger cities ≠ always worse AQI**
- **Pollution depends more on**:
    - **Industrial zones**
    - **Vehicle density**
    - **Local climate and crop burning**

Q4: How aware are Indian citizens of what AQI (Air Quality Index) means — and do they understand its health implications?

According to Chintan Environmental Research and Action Group study Air quality awareness in India is very different depending on who you ask. In cities like Delhi and nearby areas, about 7 out of 10 middle-class people know what the Air Quality Index, or AQI, means. But for people living in poorer neighborhoods, only around 1 out of 10 know about it. Younger people mostly learn about air pollution from social media, while older people rely on TV news. This means many people who live in the most polluted places don’t really understand how dangerous the air can be or how to protect themselves. So, it’s very important to have clear, easy-to-understand information about air quality for everyone, especially in poorer communities.

Q5: Several major policies and regulatory initiatives launched by the Indian government in the past five years have delivered the most **measurable impact** on improving air quality, but results have varied widely across cities and regions.

Key National Policies and Their Impact

1. National Clean Air Programme (NCAP) – Launched 2019

- Aimed for 20–30% reduction in PM2.5 and PM10 concentrations by 2024 in 131 non-attainment cities (later revised to 40% by 2026)[1][2].
- Provided over ₹9,650 crore in central funding to help cities implement targeted action plans[2].
- Measures included stricter industrial emission regulations, vehicular restrictions, waste management reforms, expanded air monitoring, and awareness campaigns.
- **Impact:**
    - Out of 131 target cities, 88 showed improvements in 2022–2023 compared to prior years, with 24 cities seeing clear falls in particulate pollution[3][4][5].
    - Cities such as Varanasi (−76%), Moradabad, Kalaburagi, Bhubaneswar, and Surat saw the biggest PM reduction, achieved by a mix of traffic controls, industrial reforms, and public transit upgrades[3][4].
    - However, a substantial number of cities (e.g., Aurangabad, Tirupati) recorded increases in pollution, often due to rapid urbanization, weak enforcement, or cross-boundary pollution[3].

2. Bharat Stage VI (BS-VI) Emission Standards – Effective April 2020

- Nationwide upgrade to vehicle emission norms equivalent to Europe's strictest standards[6][5].
- Includes tighter limits on nitrogen oxides (NOx), PM, and sulfur emissions from all new vehicles and encourages scrappage of older, more polluting vehicles.
- **Impact:**
    - Reductions in key exhaust pollutants are expected most visibly in metros like Delhi, Mumbai, and Bengaluru, where vehicle density is highest.
    - Impact is gradual and cumulative as vehicle fleets turn over[6].

3. New Industrial and Power Sector Regulations

- Expanded standards for SO_2 and NOx from coal-fired power plants[5].
- Ban on pet coke and furnace oil in NCR states; push toward PNG/biomass for factories[5].
- **Impact:**
    - Stricter enforcement in places like Varanasi, Surat, and Odisha regions has led to measurable PM reductions[3].
    - Industrial improvements are less effective where monitoring or compliance remains weak.

4. Targeted Local Actions

- Delhi NCR: Graded Response Action Plan (GRAP), dust control, hot-mix plant closures, crop stubble burning bans[75].
- Other cities: Urban forest programs (Miyawaki method), electrification of public transport, real-time public AQI displays[5].
- **Impact fluctuates** by city based on local political commitment, infrastructure, and enforcement.

Regional and City-Level Variation in Impact

Region/City	Improvements (Past 5 Years)	Key Drivers / Challenges
Varanasi	PM2.5 down by up to 76%[34]	Aggressive controls on vehicles/industry; high-level focus
Surat, Odisha, Bihar	PM2.5 drops of 40–51%[3]	Stronger industrial controls, waste management
Delhi/NCR	Marginal improvements, periods of high pollution persist[5]	Transboundary sources, crop burning, rapid vehicle growth
Mumbai	Mixed results, some rise in PM from construction and urban sprawl[36]	Implementation lag, sea breeze disperses pollutants rapidly
Aurangabad, Tirupati	Pollution levels increased[3]	Rapid urbanization, weak enforcement, lack of monitoring
Rest of India	Large variability; rural and smaller towns lack systematic air monitoring/data	Disparities in resource allocation, local challenges

Why Impacts Vary

- **Policy Execution:** Some cities efficiently use NCAP funds and enforce regulations; others do not (e.g., Visakhapatnam and Bengaluru spent less than 1% of allocated NCAP funds)[7].
- **Cross-Boundary Effects:** Many cities, especially Delhi, can only control a portion of their pollution due to inflows from adjacent states[76].
- **Urban vs Rural:** Monitoring, enforcement, and funding are concentrated in major metros; smaller cities and rural areas see less direct benefit.