# Air Quality Data Analysis



## Air Pollutants

There are two main types of air pollutants: primary and secondary. Most air pollutants come from burning fossil fuels. Some come from burning forests. Some are due to the evaporation of chemicals.

## Types of Air Pollutants

There are two basic types of pollutants in the air. They are known as primary pollutants and secondary pollutants. **Primary pollutants** enter the air directly. Some are released by natural processes, like ash from volcanoes. Most are released by human activities.

- Carbon oxides are released when fossil fuels burn.
- Nitrogen oxides form when nitrogen and oxygen combine at high temperatures. This occurs in hot exhausts from vehicles, factories, and power plants.
- Sulfur oxides are produced when sulfur and oxygen combine. This happens when coal that contains sulfur burns.
- Toxic heavy metals include mercury and lead. Mercury comes from smokestacks. Both metals have industrial uses.
- Volatile organic compounds (VOCs) are carbon compounds, such as methane. VOCs are released by many human activities. Raising livestock, for example, produces a lot of methane.
- Particulates are solid particles. These particles may be ash, dust, or even animal wastes. Many are released when fossil fuels burn.

**Secondary pollutants** form from primary pollutants. Many occur as part of **photochemical smog**. This type of smog is seen as a brown haze in the air. Photochemical smog forms when certain pollutants have a chemical reaction in the presence of sunlight. Photochemical smog consists mainly of **ozone** ($O_3$). Ozone near the ground is a pollutant. This ozone is harmful to humans and other living things. However, ozone in the stratosphere protects Earth from the Sun's harmful ultraviolet radiation.

In 2004, the National Environment Protection Council made the National Environment Protection (Air Toxics) Measure which addresses the five priority air toxics: benzene, formaldehyde, toluene, xylenes and benzo (a) pyrene (as a marker for polycyclic aromatic hydrocarbons).

# About Dataset

## Context

Air is what keeps humans alive. Monitoring it and understanding its quality is of immense importance to our well-being.

## Content

The dataset contains air quality data and AQI (Air Quality Index) at hourly and daily level of various stations across multiple cities in India.

## Cities

**Ahmedabad, Aizawl, Amaravati, Amritsar, Bengaluru, Bhopal, Brajrajnagar, Chandigarh, Chennai, Coimbatore, Delhi, Ernakulam, Gurugram, Guwahati, Hyderabad, Jaipur, Jorapokhar, Kochi, Kolkata, Lucknow, Mumbai, Patna, Shillong, Talcher, Thiruvananthapuram, Visakhapatnam**

## Acknowledgements

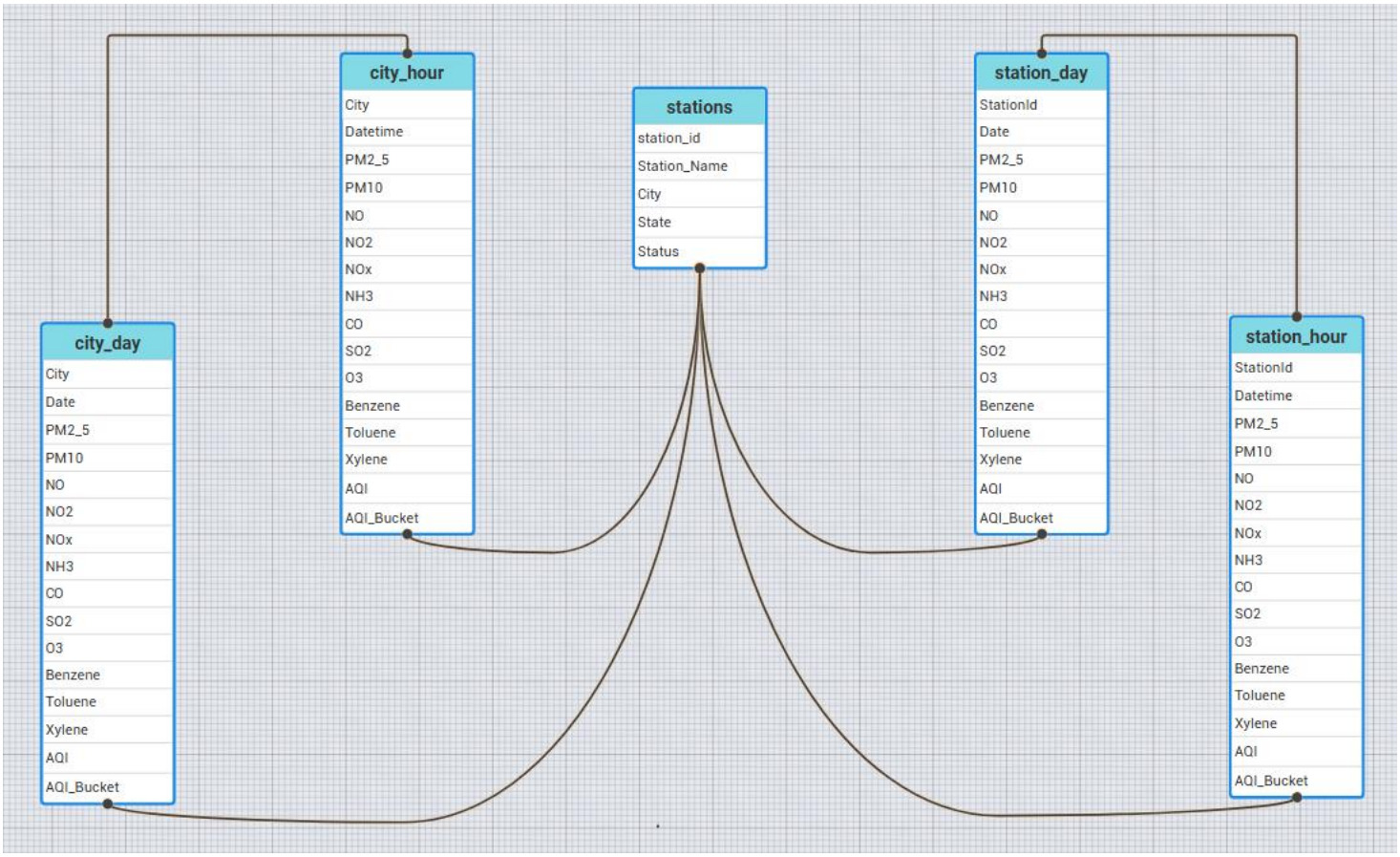
The data has been made publicly available by the Central Pollution Control Board: https://cpcb.nic.in/ which is the official portal of Government of India. They also have a real-time monitoring app: https://app.cpcbccr.com/AQI_India/

## Dataset:

https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india

The data is available in 5 different csv files:

1. City_day.csv
2. City_hour.csv
3. Station_day.csv
4. Station_hour.csv
5. Stations.csv




## Problem Statement:

Assuming you are a data analyst/ scientist you have been assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.

1. Average AQI by city (daily)

Q. For a given date range, compute daily average AQI for each city

```
select City , round(avg(AQI) ,3) as daily_average_AQI
from `Air_quality_index.city_day`
where AQI is not null
group by City
order by daily_average_AQI asc
```

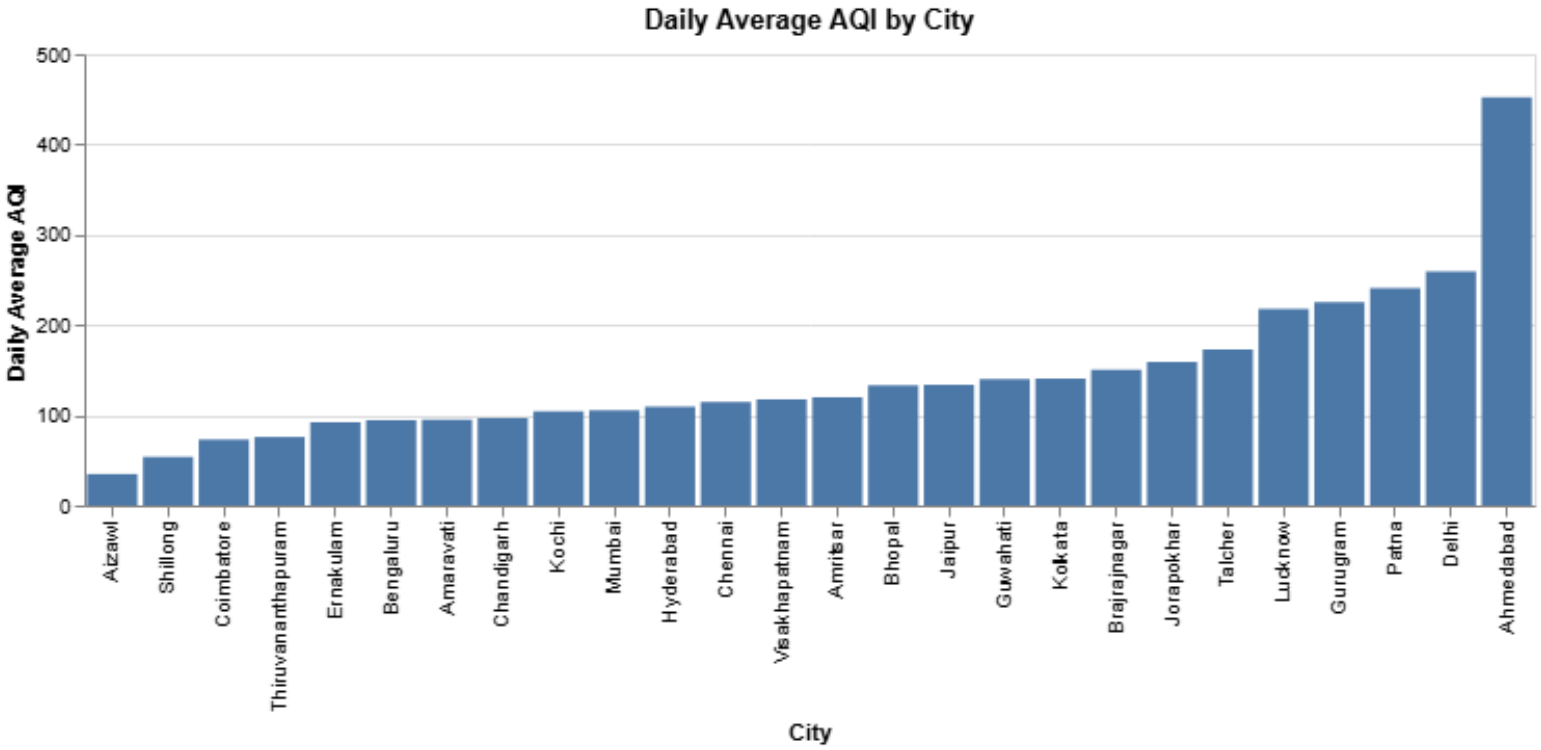
Result

Query results

Job information | Results | Visualization | JSON

| Row | City | daily_average_AQI |
|---|---|---|
| 1 | Aizawl | 34.766 |
| 2 | Shillong | 53.795 |
| 3 | Coimbatore | 73.023 |
| 4 | Thiruvananthapuram | 75.878 |
| 5 | Ernakulam | 92.359 |
| 6 | Bengaluru | 94.318 |
| 7 | Amaravati | 95.3 |
| 8 | Chandigarh | 96.498 |
| 9 | Kochi | 104.285 |
| 10 | Mumbai | 105.352 |

## Visualization




## Insight

- Ahmedabad has the highest Daily Average AQI, indicating significantly poorer air quality compared to other cities.
- Aizawl demonstrates the best air quality among the listed cities, with the lowest Daily Average AQI.
- There is a notable disparity in air quality across the cities, as evidenced by the substantial difference between the average of the five lowest and five highest AQI values.

## 2. Hourly average pollutant levels for a station

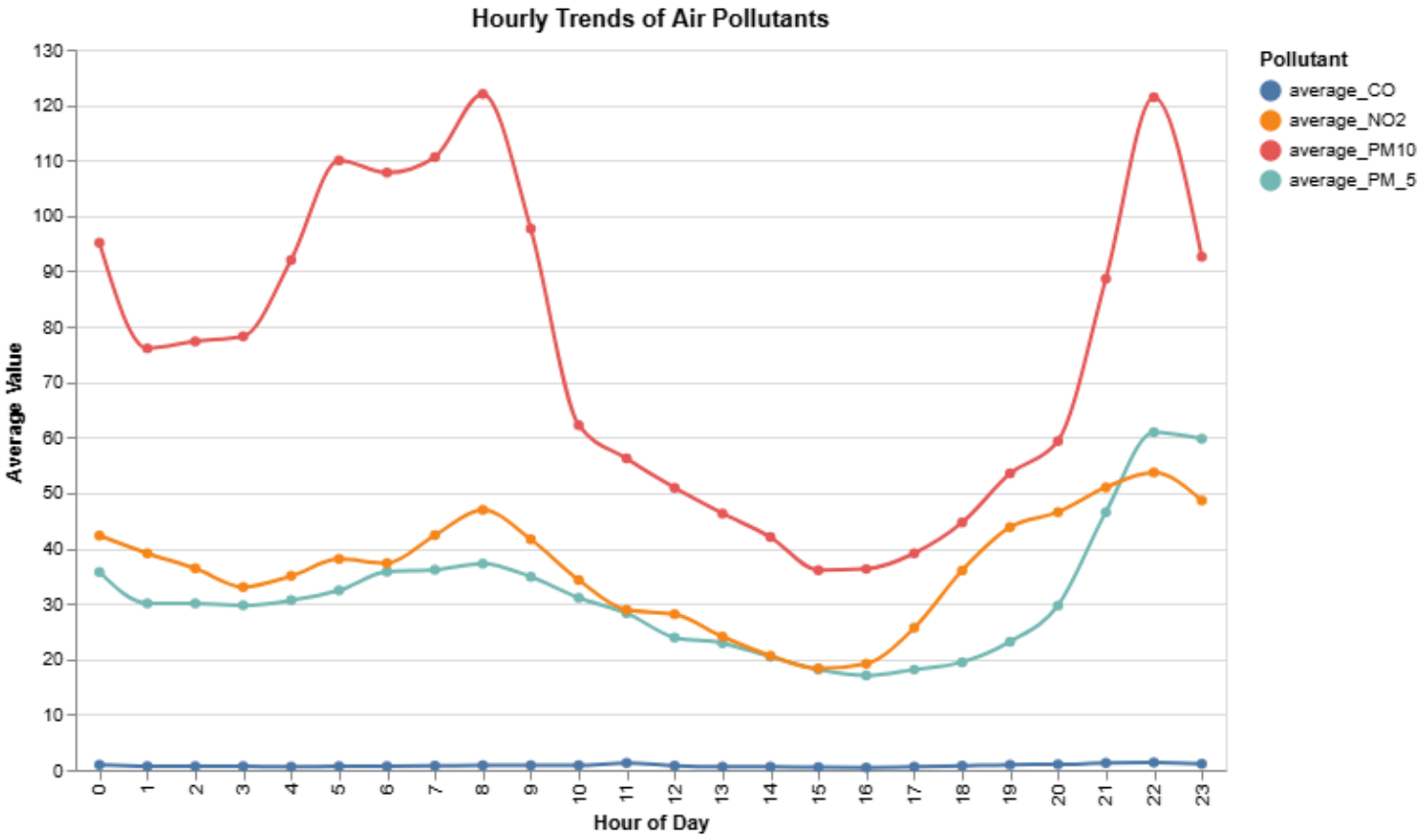
Q. For station AP001 and date range (October to december), compute average PM2_5, PM10, etc. for each hour of day (0–23).

```
select extract(hour from Datetime) as hour_of_day , round(avg(PM2_5),2) as
average_PM_5 ,
       round(avg(PM10),2) as average_PM10 ,round(avg(NO2),2) as
average_NO2 ,round(avg(NOx),2) as average_NOx,
       round(avg(NH3),2) as average_NH3 , round(avg(CO),2) as average_CO
,round(avg(SO2),2) as average_SO2 ,
       round(avg(O3),2) as average_O3 ,round(avg(Benzene),2) as
average_Benzene , round(avg(Toluene),2) as average_Toluene,
       round(avg(Xylene),2) as average_Xylene ,round(avg(AQI),2) as
average_AQI
from `Air_quality_index.station_hour`
where  extract(date from Datetime) between '2019-10-01' and '2019-10-
31'  and StationId = 'AP005'
group by extract(hour from Datetime)
```

## Result


| Row | hour_of_day | average_PM_5 | average_PM10 | average_NO2 | average_NOx | average_NH3 | average_CO | average_SO2 | average_O3 | average_Benzene | average_Toluene | average_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 35.74 | 95.12 | 42.28 | 39.2 | 8.7 | 0.9 | 24.99 | 22.37 | 4.38 | 13.63 | |
| 2 | 1 | 30.06 | 76.1 | 39.1 | 34.78 | 8.55 | 0.65 | 24.73 | 22.04 | 3.39 | 11.56 | |
| 3 | 2 | 30.02 | 77.35 | 36.37 | 31.9 | 8.19 | 0.65 | 24.11 | 23.26 | 3.51 | 10.91 | |
| 4 | 3 | 29.71 | 78.25 | 33.03 | 29.27 | 10.06 | 0.64 | 23.14 | 23.6 | 3.11 | 10.46 | |
| 5 | 4 | 30.61 | 92.03 | 35.0 | 32.43 | 13.18 | 0.58 | 26.2 | 19.73 | 2.86 | 11.04 | |
| 6 | 5 | 32.39 | 110.0 | 38.04 | 40.23 | 9.75 | 0.67 | 32.01 | 20.68 | 3.1 | 11.13 | |
| 7 | 6 | 35.76 | 107.77 | 37.31 | 42.38 | 8.11 | 0.67 | 23.67 | 20.99 | 3.6 | 11.79 | |
| 8 | 7 | 36.11 | 110.57 | 42.37 | 49.27 | 8.3 | 0.73 | 22.56 | 19.88 | 3.86 | 12.11 | |
| 9 | 8 | 37.25 | 121.99 | 46.96 | 45.19 | 8.74 | 0.86 | 22.98 | 21.44 | 4.41 | 12.42 | |

## Visualization




## Insight

- The dataset covers a complete 24-hour period, representing a full daily cycle.
- The AQI (Air Quality Index) values vary roughly between **~86 and ~99+**

3. Station-level worst day

Q. For each station, find the day (from station_day) with the highest daily AQI, and report that day and AQI.

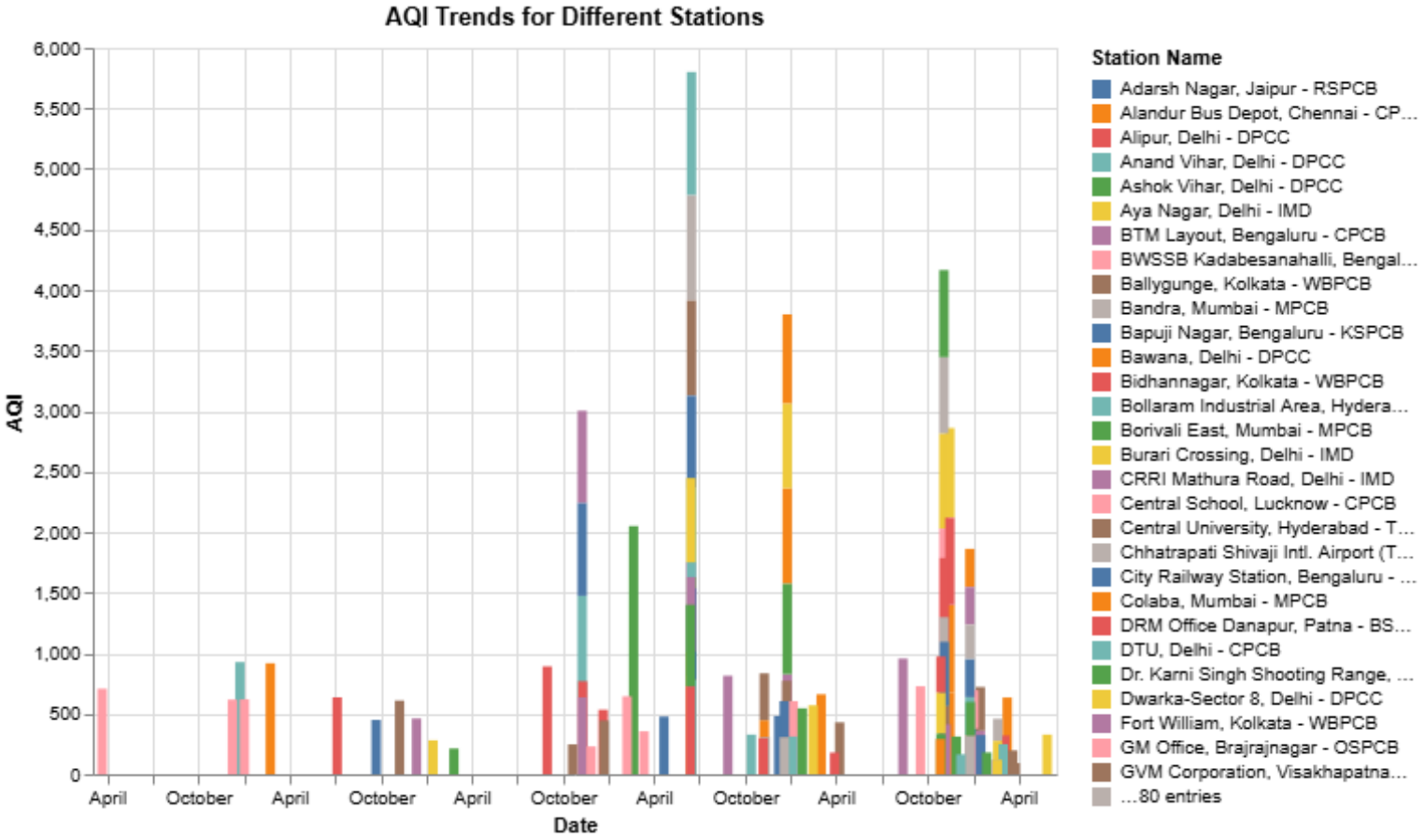
```
select Station_Name ,Date ,AQI
from (
       select Station_Name,Date ,A.AQI ,dense_rank() over (partition
       by A.StationId order by A.AQI desc) as highest_of_day
      from `Air_quality_index.station_day` A
      inner join `Air_quality_index.stations` B
      on A.StationId = B.station_id ) a
where highest_of_day = 1 and AQI is not null
order by AQI desc
```

Query results

Job information | Results | Visualization | JSON | Execution details | Execution graph

| Row | Station_Name | Date | AQI |
|---|---|---|---|
| 1 | Maninagar, Ahmedabad - GPCB | 2018-02-19 | 2049 |
| 2 | DTU, Delhi - CPCB | 2018-06-15 | 1019 |
| 3 | Sirifort, Delhi - CPCB | 2018-06-14 | 1012 |
| 4 | Railway Colony, Guwahati - APCB | 2019-08-13 | 956 |
| 5 | Anand Vihar, Delhi - DPCC | 2015-12-23 | 927 |
| 6 | Sanathnagar, Hyderabad - TSPCB | 2016-02-22 | 917 |
| 7 | Vikas Sadan, Gurugram - HSPCB | 2017-08-30 | 891 |
| 8 | Golden Temple, Amritsar - PPCB | 2018-06-15 | 869 |
| 9 | R K Puram, Delhi - DPCC | 2017-11-08 | 838 |

## Result




## Visualization

## Insight

- AQI peaked at 2049 in Maninagar, Ahmedabad (GPCB) on 2018-2-19, representing the maximum observed value.
- The lowest AQI recorded was 92 in Sikulpuikawn, Aizawl (Mizoram PCB) on 2020-3-25, indicating a significant difference from the maximum.

4. Most frequent AQI bucket per station

Q. For each station, count how many days fell into each AQI_Bucket (Good, Moderate, etc.), and show the bucket which appears most.

```
select Station_Name , AQI_Bucket , count(*) as count_AQI_bucket
from `Air_quality_index.station_day` A
join `Air_quality_index.stations` B
on A.StationId = B.station_id
where AQI_Bucket is not null
group by Station_Name , AQI_Bucket
order by count_AQI_bucket desc
```

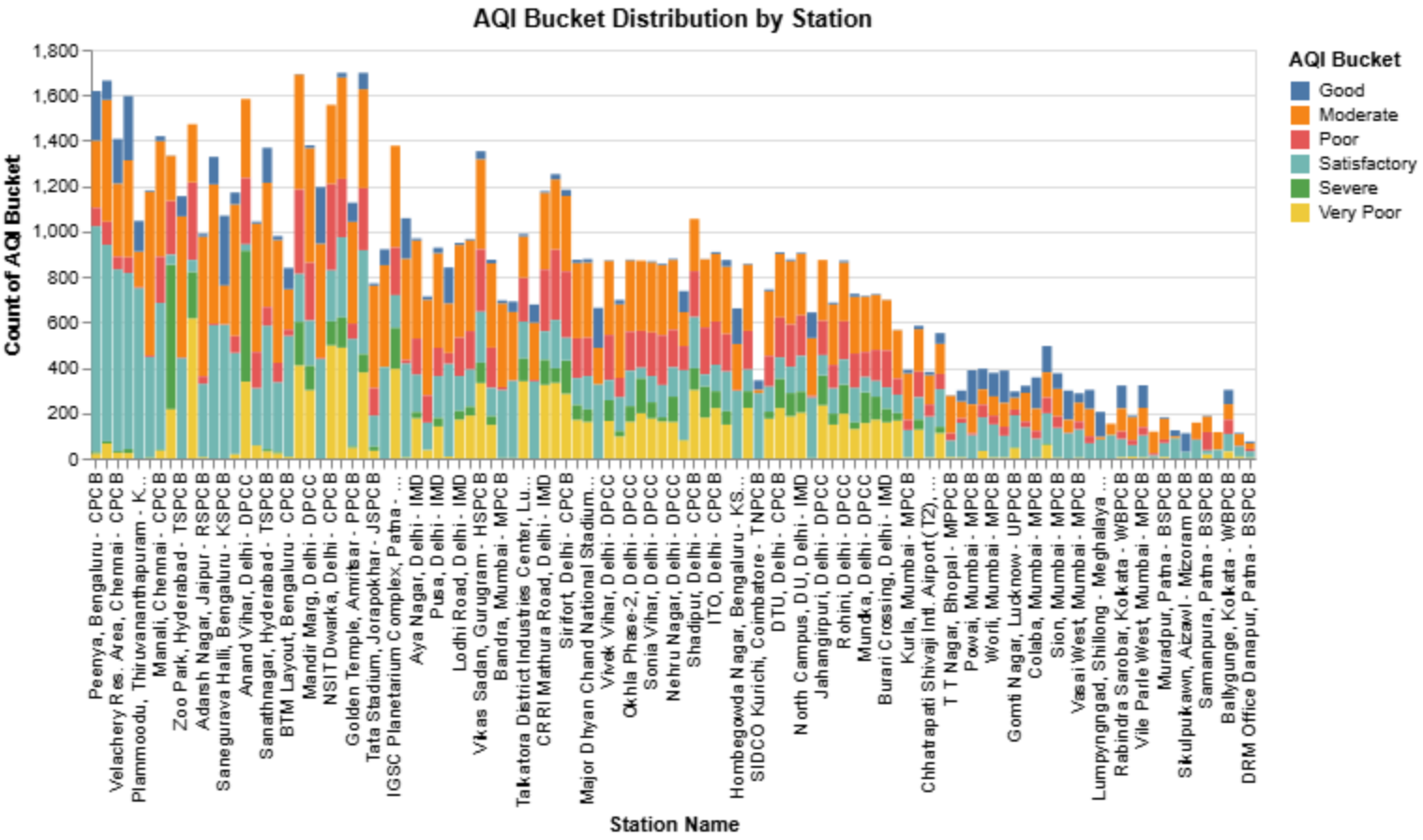
Result

Query results

Job information | Results | Visualization | JSON | Execution details

| Row | Station_Name | AQI_Bucket | count_AQI_bucket |
|---|---|---|---|
| 1 | Peenya, Bengaluru - CPCB | Satisfactory | 994 |
| 2 | Alandur Bus Depot, Chennai - C... | Satisfactory | 866 |
| 3 | Velachery Res. Area, Chennai - ... | Satisfactory | 800 |
| 4 | BWSSB Kadabesanahalli, Benga... | Satisfactory | 776 |
| 5 | Plammoodu, Thiruvananthapura... | Satisfactory | 750 |
| 6 | City Railway Station, Bengaluru -... | Moderate | 721 |
| 7 | Manali, Chennai - CPCB | Satisfactory | 647 |
| 8 | Maninagar, Ahmedabad - GPCB | Severe | 638 |
| 9 | Zoo Park, Hyderabad - TSPCB | Moderate | 620 |

## Visualization




## Insight

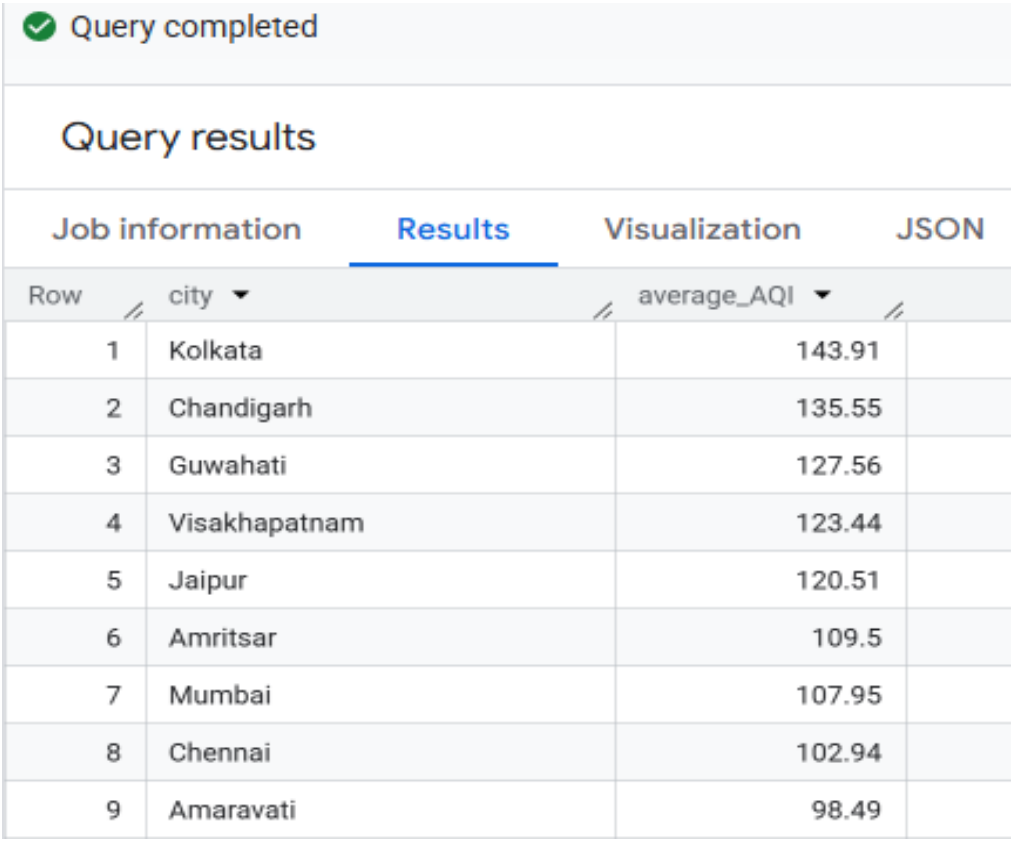
- Moderate AQI is most frequent, accounting for 33.8% of the total count, while Severe AQI is the least frequent at 5.98%.
- Central School and Lalbagh in Lucknow each contribute 1.95% to the total AQI Bucket count; Satisfactory AQI is 26.97% of Central School's total, while Very Poor AQI is 28.74% of Lalbagh's total.
- Peenya, Bengaluru - CPCB station shows a relatively high count of Satisfactory AQI, representing 1.14% of the total count and 4.21% of all Satisfactory readings.

## 5. Top 10 most polluted cities (yearly)

Q. For a 2019 year, rank the top 10 cities by their annual average AQI.

```
select city, average_AQI
from ( select city, round(avg(AQI), 2) as average_AQI, dense_rank() over
(order by AVG(AQI) desc) as rnk
  from `Air_quality_index.city_day`
  where Date between '2019-01-01' and '2019-12-31'
  group by  city ) a
where rnk >= 10
```

## Result



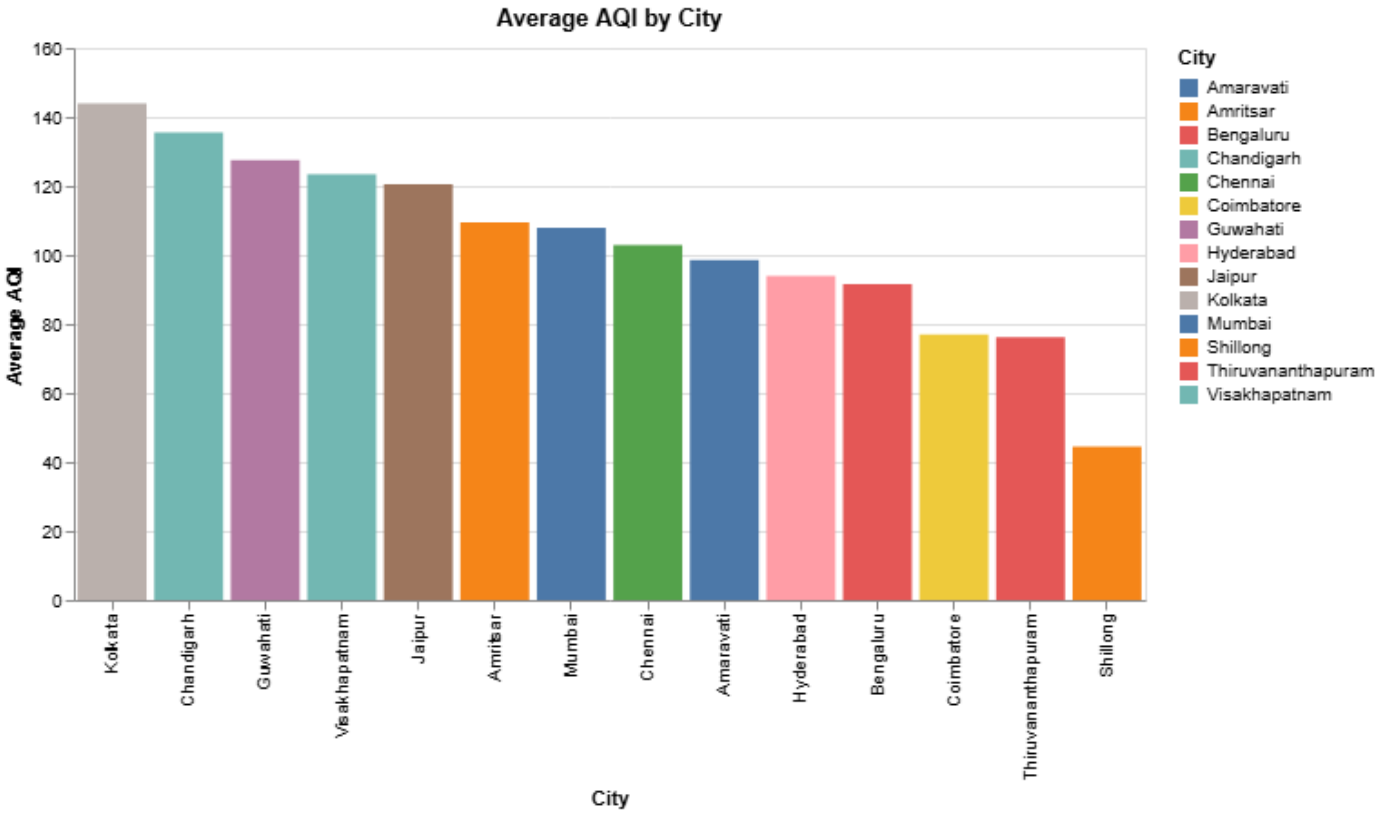
Query completed

Query results

Job information | Results | Visualization | JSON

| Row | city | average_AQI |
|---|---|---|
| 1 | Kolkata | 143.91 |
| 2 | Chandigarh | 135.55 |
| 3 | Guwahati | 127.56 |
| 4 | Visakhapatnam | 123.44 |
| 5 | Jaipur | 120.51 |
| 6 | Amritsar | 109.5 |
| 7 | Mumbai | 107.95 |
| 8 | Chennai | 102.94 |
| 9 | Amaravati | 98.49 |

## Visualization




## Insight

- Shillong exhibits the lowest average AQI, suggesting the best air quality in the dataset.
- A general trend of decreasing average AQI is observed moving southward, with southern cities reporting lower pollution levels.
- The top 7 cities, including Kolkata and Mumbai, have a higher average AQI, implying relatively elevated pollution in these metropolitan areas.
- Kolkata has the highest average AQI, indicating the poorest air quality among the cities analyzed.

## 6. Monthly trend of PM2_5 for a city

For a (bengaluru) city , compute the monthly average PM2_5 for all months in a given year, in order, to see the trend.

```
select year , month ,round( average_PM2_5 - lag(average_PM2_5) over (order
by year , month) ,2) as monthly_trend
from (
      select extract(year from date ) as year , extract(month from date)
as month , round(avg(PM2_5),2) as average_PM2_5
      from `Air_quality_index.city_day`
       where City = 'Bengaluru' and PM2_5 is not null
      group by extract(year from date) , extract(month from date) )a
where average_PM2_5 is not null
```

## Result

Query results

Job information | Results | Visualization | JSON | Execution details

| Row | year | month | monthly_trend |
|---|---|---|---|
| 1 | 2015 | 3 | null |
| 2 | 2015 | 4 | -18.0 |
| 3 | 2015 | 5 | 4.88 |
| 4 | 2015 | 6 | -15.25 |
| 5 | 2015 | 7 | 2.72 |
| 6 | 2015 | 8 | 0.1 |
| 7 | 2015 | 9 | -0.89 |
| 8 | 2015 | 10 | 11.2 |
| 9 | 2015 | 11 | -0.66 |

## Visualizaton


## Insight

- December 2017 saw the largest monthly increase at 28.84%, while January 2018 experienced the biggest decline at -28.51%.
- A seasonal pattern exists where December typically shows positive spikes, contrasted by declines often observed in January.
- April and June from 2015 to 2020 generally showed negative monthly trends.
- October 2016 had a significantly higher monthly trend (18.75%) compared to October 2015 (11.2%).

## 7. Station with most active status

Q. Count readings (days) per station, and list the top few stations with the most data (highest count).

```
select s.Station_id, s.Station_Name, s.City, s.State, s.Status,
count(*) as reading_count
from `Air_quality_index.stations` as s
join `Air_quality_index.station_day` as r
on s.Station_id = r.StationId
group by  s.Station_id, s.Station_Name, s.City, s.State, s.Status
order by reading_count desc
limit 10
```

## Result

Query results

Job information | Results | Visualization | JSON | Execution details | Execution graph

| Row | Station_id | Station_Name | City | State | Status | reading_count |
|---|---|---|---|---|---|---|
| 1 | KA009 | Peenya, Bengaluru - CPCB | Bengaluru | Karnataka | Active | 2009 |
| 2 | DL008 | DTU, Delhi - CPCB | Delhi | Delhi | Active | 2009 |
| 3 | TN001 | Alandur Bus Depot, Chennai - C... | Chennai | Tamil Nadu | Active | 2009 |
| 4 | MH005 | Bandra, Mumbai - MPCB | Mumbai | Maharashtra | Active | 2009 |
| 5 | DL013 | IHBAS, Dilshad Garden, Delhi - C... | Delhi | Delhi | Active | 2009 |
| 6 | DL007 | CRRI Mathura Road, Delhi - IMD | Delhi | Delhi | Active | 2009 |
| 7 | GJ001 | Maninagar, Ahmedabad - GPCB | Ahmedabad | Gujarat | Active | 2009 |
| 8 | KA003 | BWSSB Kadabesanahalli, Benga... | Bengaluru | Karnataka | Active | 2009 |
| 9 | DL033 | Shadipur, Delhi - CPCB | Delhi | Delhi | Active | 2009 |

## insight

- The dataset includes data from 10 active air quality monitoring stations distributed across 5 cities and 5 states, with each station having 2009 readings.
- Delhi is the most represented city with 5 stations, comprising 50% of the total stations in the dataset.
- Delhi and Karnataka are the most represented states, each accounting for 50% of the total stations.
- All listed stations are active and possess an equal number of readings, indicating a uniform data collection timeframe across all locations

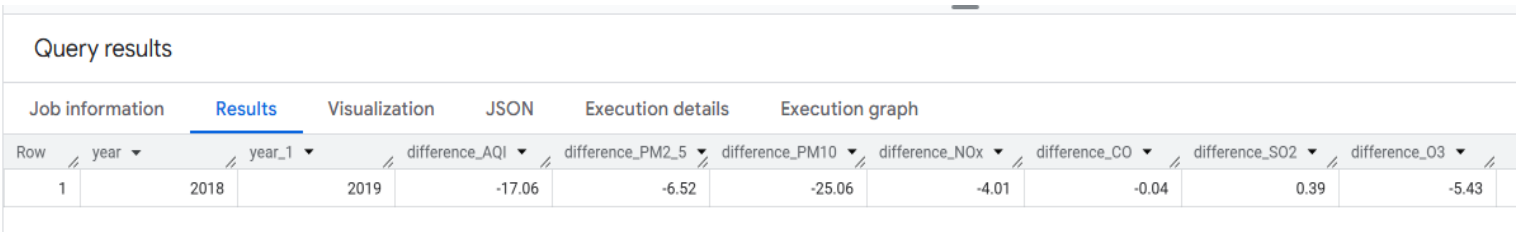
## 8. Before vs after policy

Q. Suppose a regulation began on 2019-01-01. For delhi city, compare average AQI and pollutant levels before vs after in given time windows.

```
select A.year,B.year , round((average_AQI_2019 - average_AQI_2018),2) as
difference_AQI , round((average_PM2_5_2019 - average_PM2_5_2018),2) as
difference_PM2_5,round(( average_PM10_2019 - average_PM10_2018 ),2) as
difference_PM10 ,round((average_NOx_2019 - average_NOx_2018 ),2) as
difference_NOx ,round((average_CO_2019 - average_CO_2018 ),2) as
difference_CO ,round((average_SO2_2019 - average_SO2_2018 ),2) as
difference_SO2 ,round((average_O3_2019 - average_O3_2018 ),2) as
difference_O3

from
     (select City , extract(year from date) as year,round(avg(AQI),2) as
average_AQI_2018 , round(avg(PM2_5),2) as average_PM2_5_2018 ,
round(avg(PM10),2) as average_PM10_2018 ,round(avg(NOx),2) as
average_NOx_2018 , round(avg(CO),2) as average_CO_2018 , round(avg(SO2),2)
as average_SO2_2018 ,round(avg(O3),2) as average_O3_2018
      from `Air_quality_index.city_day`
      where City = 'Delhi' and
            date between '2018-01-01’ and'2018 -12-31'
      group by City , extract(year from date) ) A
      join
     (select City , extract(year from date) as   year,round(avg(AQI),2) as
average_AQI_2019 , round(avg(PM2_5),2) as average_PM2_5_2019 ,
round(avg(PM10),2) as average_PM10_2019 ,round(avg(NOx),2) as
average_NOx_2019 , round(avg(CO),2) as average_CO_2019 , round(avg(SO2),2)
as average_SO2_2019 ,round(avg(O3),2) as average_O3_2019
      from `Air_quality_index.city_day`
      where City = 'Delhi' and
       date between '2019-01-01' and '201912-31'
      group by City ,extract(year from date) ) B
on A.City = B.City
```

## Result



Query results

Job information | Results | Visualization | JSON | Execution details | Execution graph

| Row | year | year_1 | difference_AQI | difference_PM2_5 | difference_PM10 | difference_NOx | difference_CO | difference_SO2 | difference_O3 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2018 | 2019 | -17.06 | -6.52 | -25.06 | -4.01 | -0.04 | 0.39 | -5.43 |

9. Day-of-week effect

Q. In a city(chennai), compute average AQI by day of week (Monday, Tuesday, …) to see which days typically have worse air.

```
select date , format_datetime('%A' , date) as day_of_week , average_AQI
from (
     select date , avg(AQI) as average_AQI
     from `Air_quality_index.city_day`
```

```
    where City = 'Chennai' and AQI is not null
    group by date ) a
order by average_AQI desc
```

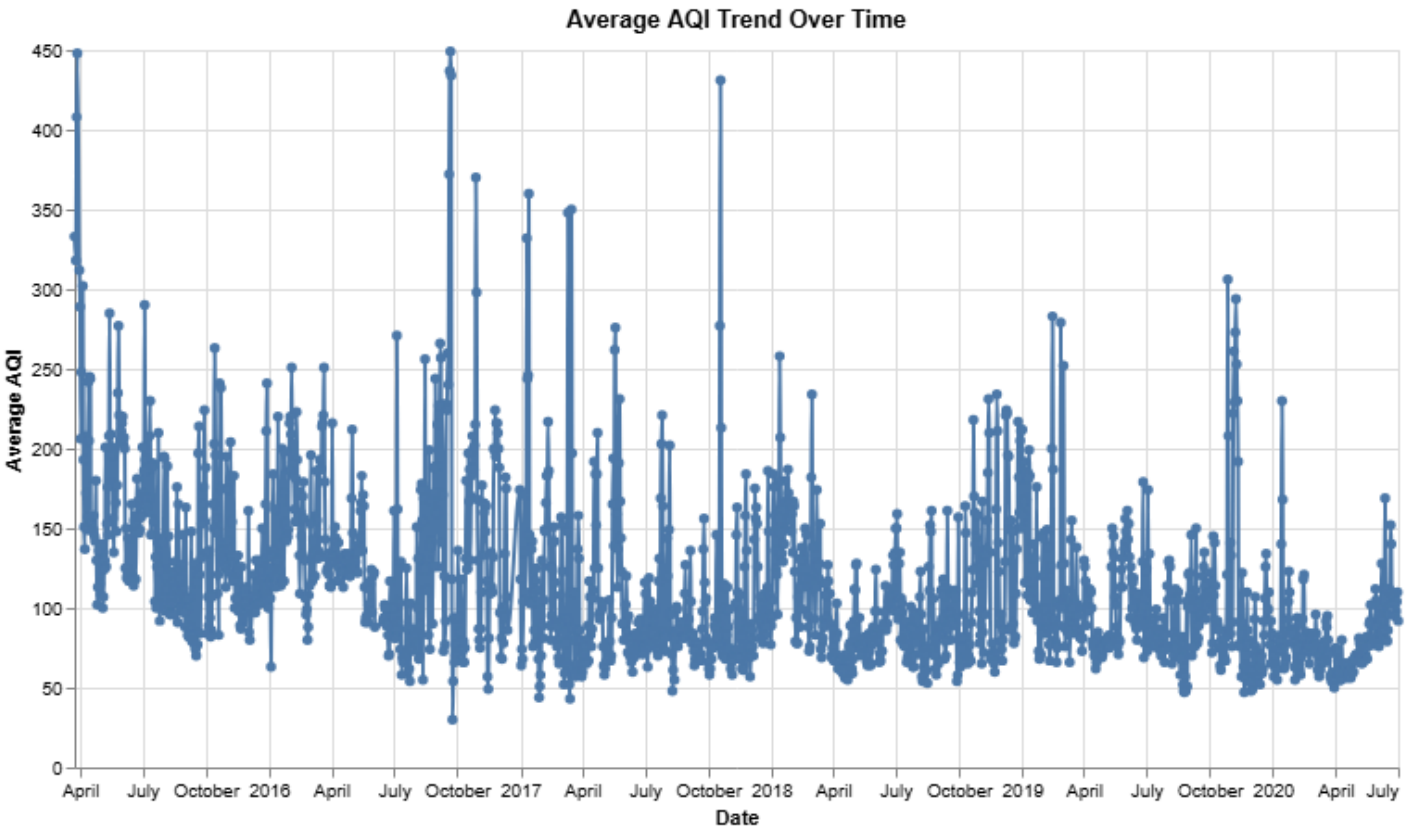
## Result

Query results

Job information | Results | Visualization | JSON | Execution details

| Row | date | day_of_week | average_AQI |
|---|---|---|---|
| 1 | 2016-09-21 | Wednesday | 449.0 |
| 2 | 2015-03-28 | Saturday | 448.0 |
| 3 | 2016-09-20 | Tuesday | 437.0 |
| 4 | 2016-09-22 | Thursday | 434.0 |
| 5 | 2017-10-19 | Thursday | 431.0 |
| 6 | 2015-03-27 | Friday | 408.0 |
| 7 | 2016-09-19 | Monday | 372.0 |
| 8 | 2016-10-28 | Friday | 370.0 |
| 9 | 2017-01-13 | Friday | 360.0 |

## Visaulization




## Insight

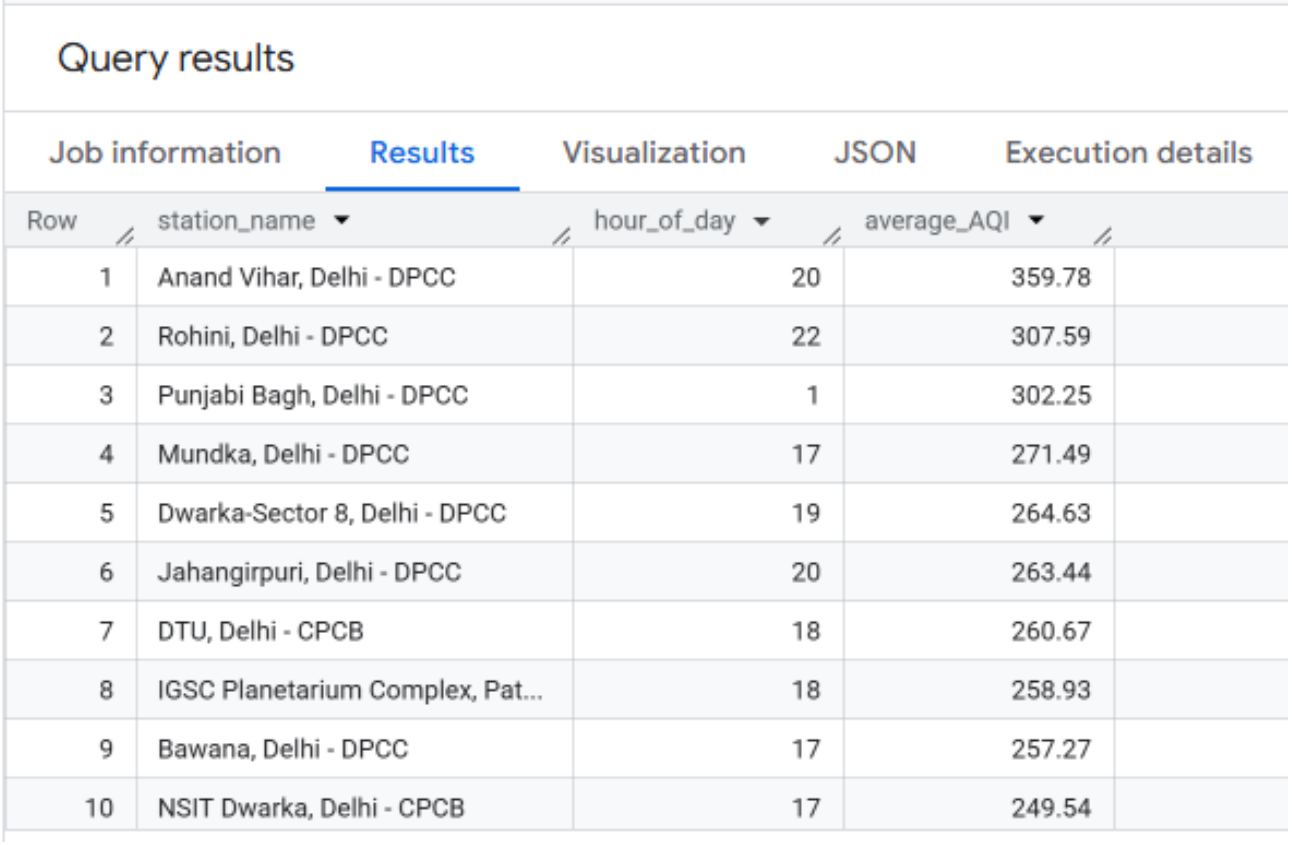
- Overall, the average AQI decreased by 72.37% from 333 on March 24, 2015, to 92 on July 1, 2020.
- A significant decrease of 93.32% in average AQI was observed from 449 on September 21, 2016, to a minimum of 30 on September 24, 2016.
- The average AQI decreased by 68.17% on Wednesdays and 66.04% on Thursdays during the observed period.

## 10. Hourly peaks

For each station, find the hour (0–23) at which average AQI is highest.

```
select station_name , hour_of_day , round(average_AQI,2) as average_AQI
from (
      select station_id,Station_Name , extract(hour from Datetime) as
hour_of_day , avg(AQI) as average_AQI ,
      dense_rank() over ( partition by station_id order by avg(AQI) desc)
as rnk
      from `Air_quality_index.station_hour` A
      join `Air_quality_index.stations` B
      on A.StationId = B.station_id
      where AQI is not null
      group by station_id , Station_Name , extract(hour from Datetime)
      order by average_AQI desc ) a
  where rnk = 1
  order by average_AQI desc
```

## Result

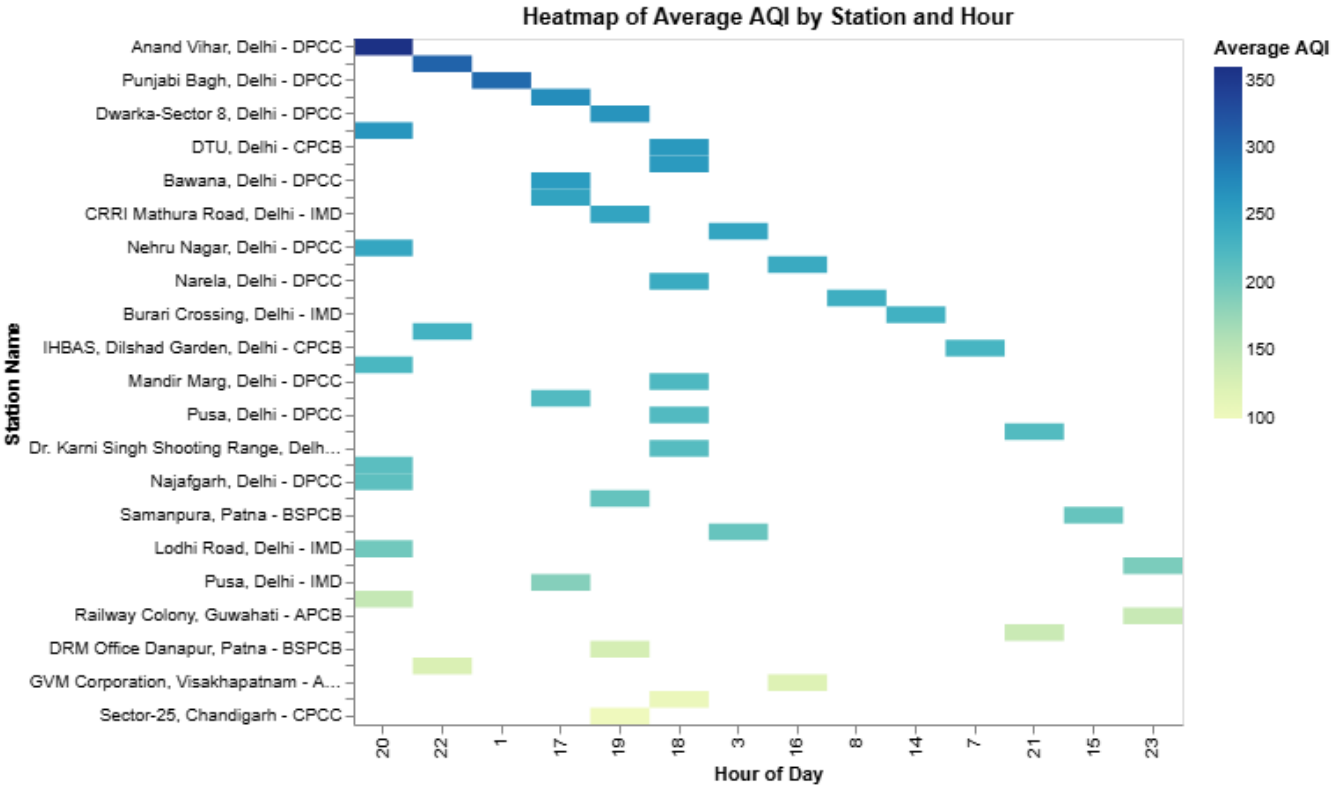


Query results

Job information | Results | Visualization | JSON | Execution details

| Row | station_name | hour_of_day | average_AQI |
|---|---|---|---|
| 1 | Anand Vihar, Delhi - DPCC | 20 | 359.78 |
| 2 | Rohini, Delhi - DPCC | 22 | 307.59 |
| 3 | Punjabi Bagh, Delhi - DPCC | 1 | 302.25 |
| 4 | Mundka, Delhi - DPCC | 17 | 271.49 |
| 5 | Dwarka-Sector 8, Delhi - DPCC | 19 | 264.63 |
| 6 | Jahangirpuri, Delhi - DPCC | 20 | 263.44 |
| 7 | DTU, Delhi - CPCB | 18 | 260.67 |
| 8 | IGSC Planetarium Complex, Pat... | 18 | 258.93 |
| 9 | Bawana, Delhi - DPCC | 17 | 257.27 |
| 10 | NSIT Dwarka, Delhi - CPCB | 17 | 249.54 |

## Visualization




## Insight

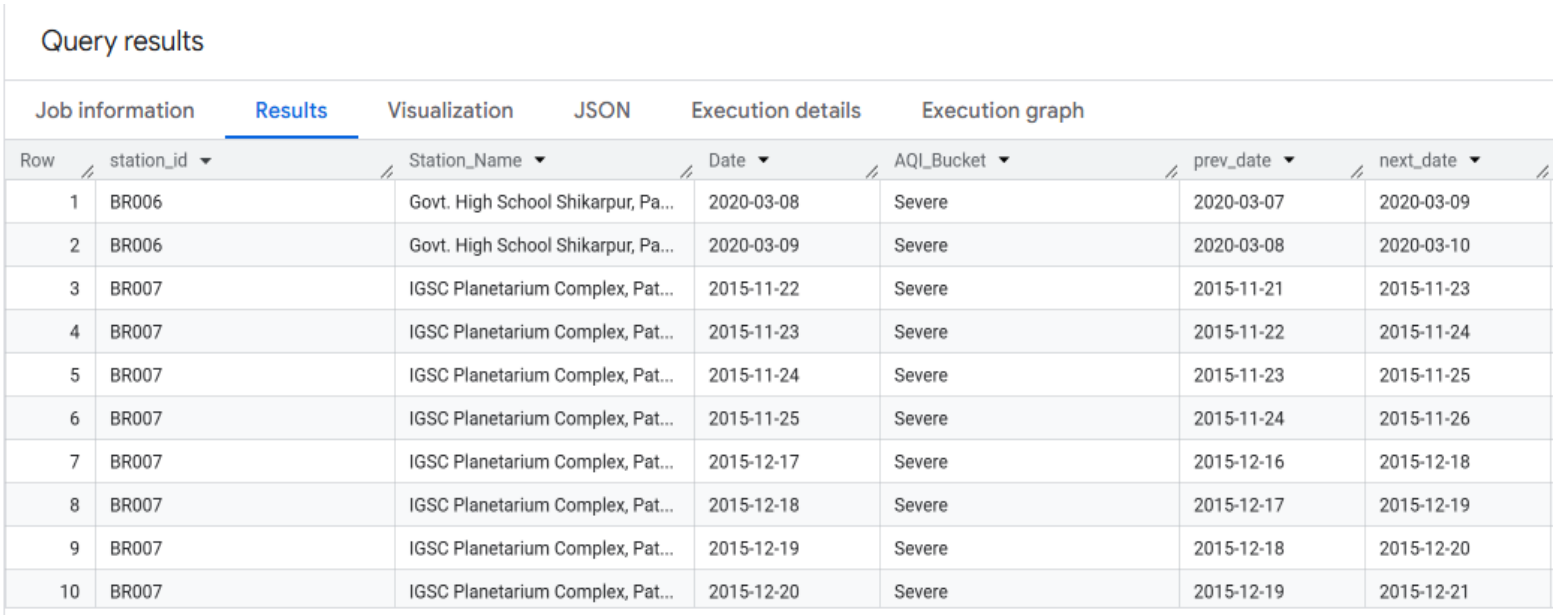
- Anand Vihar, Delhi, had the highest average AQI among the stations recorded.
- Sector-25, Chandigarh, registered the lowest average AQI in the dataset.
- Data reporting frequency peaks around Hour 20 across the stations.
- Patna stations generally show lower AQI values than Delhi stations in this dataset

## 11. Consecutive high pollution days

Q. For a station, find sequences of days where AQI was “Severe” (or above a threshold) for 3 consecutive days.

```
select station_id ,Station_Name , Date , AQI_Bucket , prev_date ,
next_date
from (
select B.station_id , B.Station_Name , A.Date , A.AQI_Bucket ,
        lag(A.Date) over (partition by A.StationId order by A.Date)
as  prev_date ,
        lead(A.Date) over (partition by A.StationId order by A.Date) as
next_date,
        lag(AQI_Bucket) over (partition by StationId order by A.Date) as
prev_bucket,
        lead(AQI_Bucket) over (partition by StationId order by A.Date) as
next_bucket
From `Air_quality_index.station_day` A
join `Air_quality_index.stations` B
on A.StationId = B.station_id ) a
where AQI_Bucket = 'Severe' and prev_bucket = 'Severe' and next_bucket =
'Severe' and
      datetime_diff(Date , prev_date,day) = 1 and
      datetime_diff(next_date ,Date , day) = 1
```

## Result

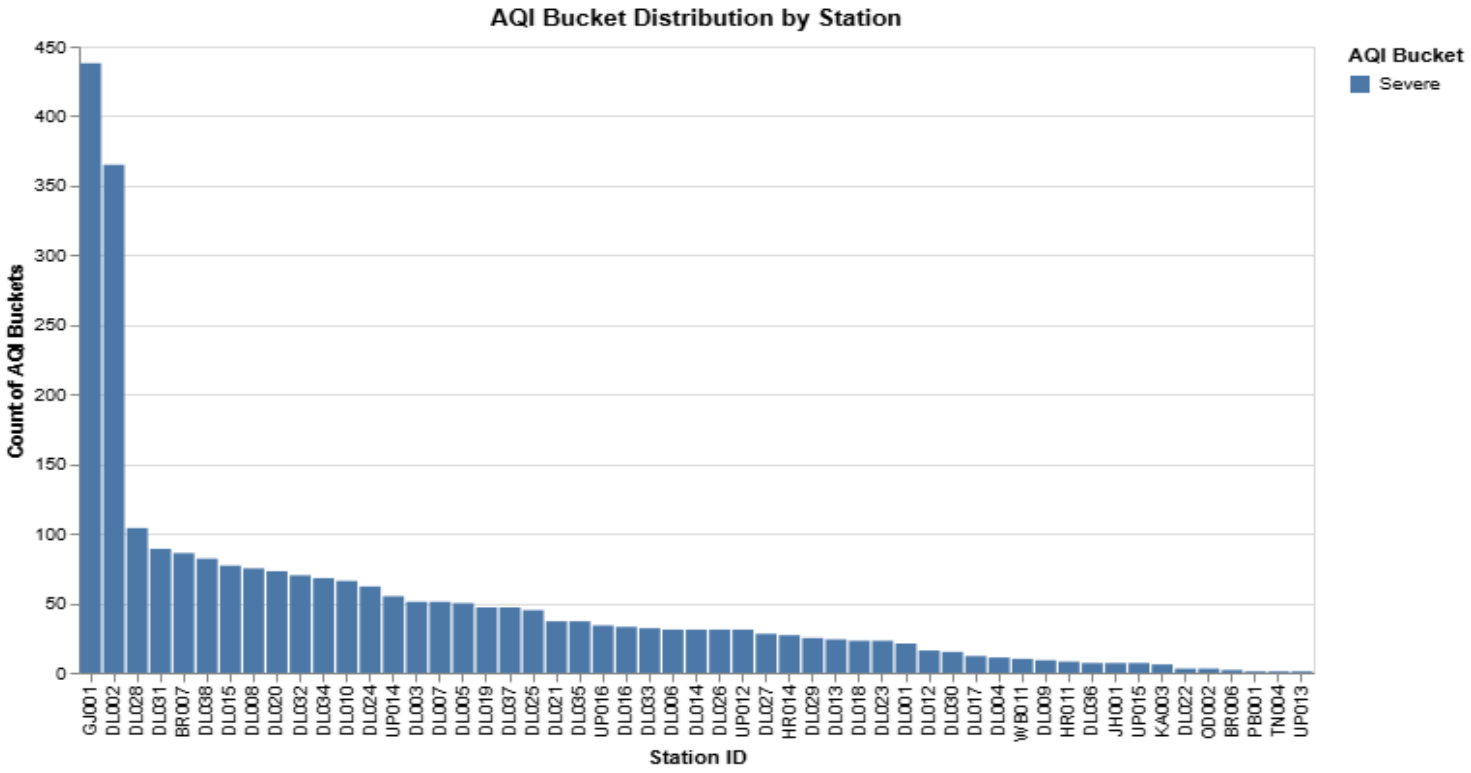


Query results

Job information | Results | Visualization | JSON | Execution details | Execution graph

| Row | station_id | Station_Name | Date | AQI_Bucket | prev_date | next_date |
|---|---|---|---|---|---|---|
| 1 | BR006 | Govt. High School Shikarpur, Pa... | 2020-03-08 | Severe | 2020-03-07 | 2020-03-09 |
| 2 | BR006 | Govt. High School Shikarpur, Pa... | 2020-03-09 | Severe | 2020-03-08 | 2020-03-10 |
| 3 | BR007 | IGSC Planetarium Complex, Pat... | 2015-11-22 | Severe | 2015-11-21 | 2015-11-23 |
| 4 | BR007 | IGSC Planetarium Complex, Pat... | 2015-11-23 | Severe | 2015-11-22 | 2015-11-24 |
| 5 | BR007 | IGSC Planetarium Complex, Pat... | 2015-11-24 | Severe | 2015-11-23 | 2015-11-25 |
| 6 | BR007 | IGSC Planetarium Complex, Pat... | 2015-11-25 | Severe | 2015-11-24 | 2015-11-26 |
| 7 | BR007 | IGSC Planetarium Complex, Pat... | 2015-12-17 | Severe | 2015-12-16 | 2015-12-18 |
| 8 | BR007 | IGSC Planetarium Complex, Pat... | 2015-12-18 | Severe | 2015-12-17 | 2015-12-19 |
| 9 | BR007 | IGSC Planetarium Complex, Pat... | 2015-12-19 | Severe | 2015-12-18 | 2015-12-20 |
| 10 | BR007 | IGSC Planetarium Complex, Pat... | 2015-12-20 | Severe | 2015-12-19 | 2015-12-21 |

## Visualization




## Insight

- Data from the Maninagar, Ahmedabad - GPCB station represents a significant portion (16.92%) of the overall dataset.
- The date 2019-11-14 is the most frequently recorded date in the dataset, accounting for 1.55% of the total entries.
- The dataset covers a multi-year period from early 2015 to early 2020, but only reflects instances of "Severe" air quality.
- The Air Quality Index (AQI) was consistently "Severe" across all reporting stations throughout the recorded period.

## 12. Transition analysis

Q. For station or city: how many times did AQI category go from "Moderate" to "Severe" (or jump two categories) day to day.

```
select City , count(1) as moderate_to_severe_count
from (
      select City , Date , AQI_Bucket , lead(Date) over (partition
by City order by Date) as next_date ,     lead(AQI_Bucket) over
(partition by City order by Date ) as    AQI_bucket_next
      from `Air_quality_index.city_day` ) A
 where AQI_Bucket = 'Moderate' and AQI_Bucket_next = 'Severe'and
      date_diff(next_date ,Date ,DAY ) = 1

group by City
```

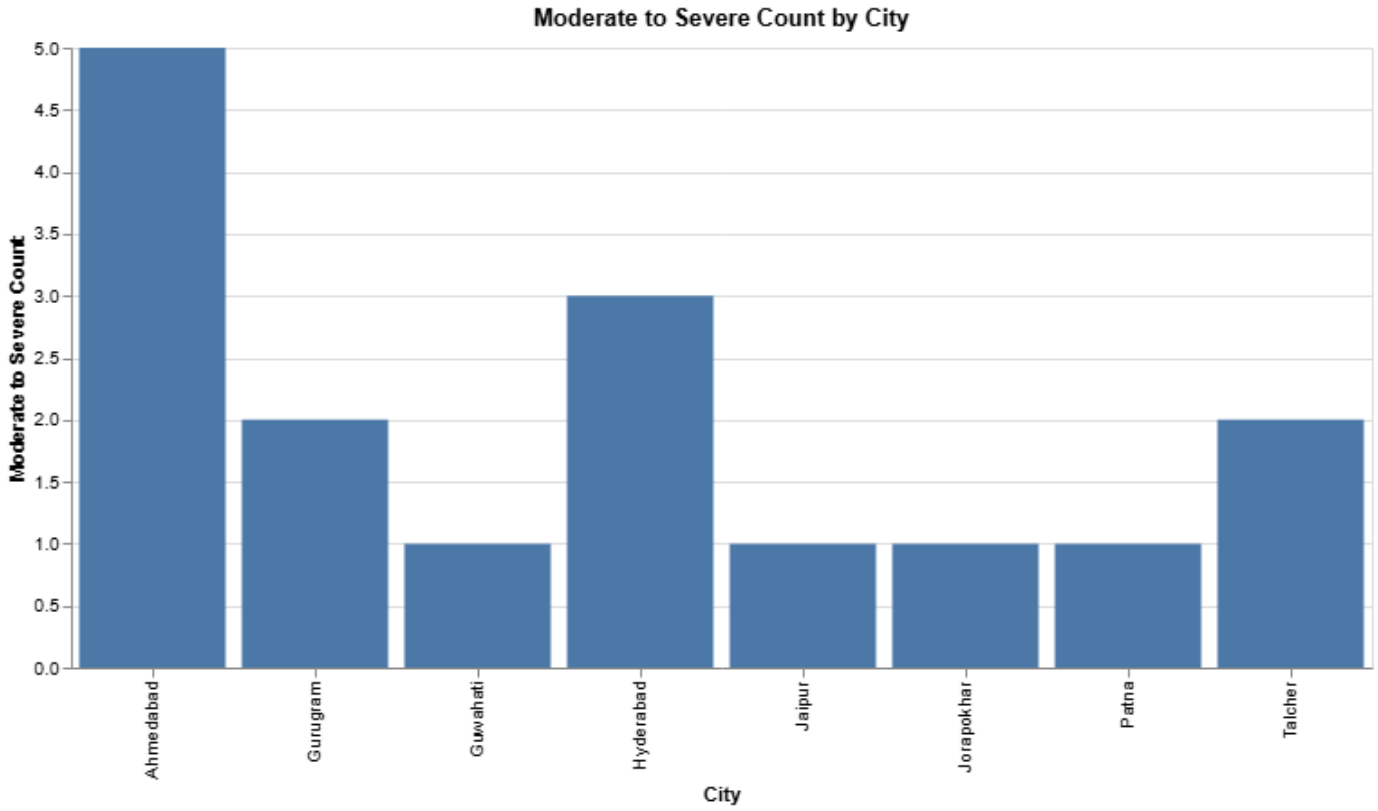
## Result

Query results

Job information | Results | Visualization | JSON

| Row | City | moderate_to_sev... |
|---|---|---|
| 1 | Ahmedabad | 5 |
| 2 | Gurugram | 2 |
| 3 | Guwahati | 1 |
| 4 | Hyderabad | 3 |
| 5 | Jaipur | 1 |
| 6 | Jorapokhar | 1 |
| 7 | Patna | 1 |
| 8 | Talcher | 2 |

## Visualization




## Insight

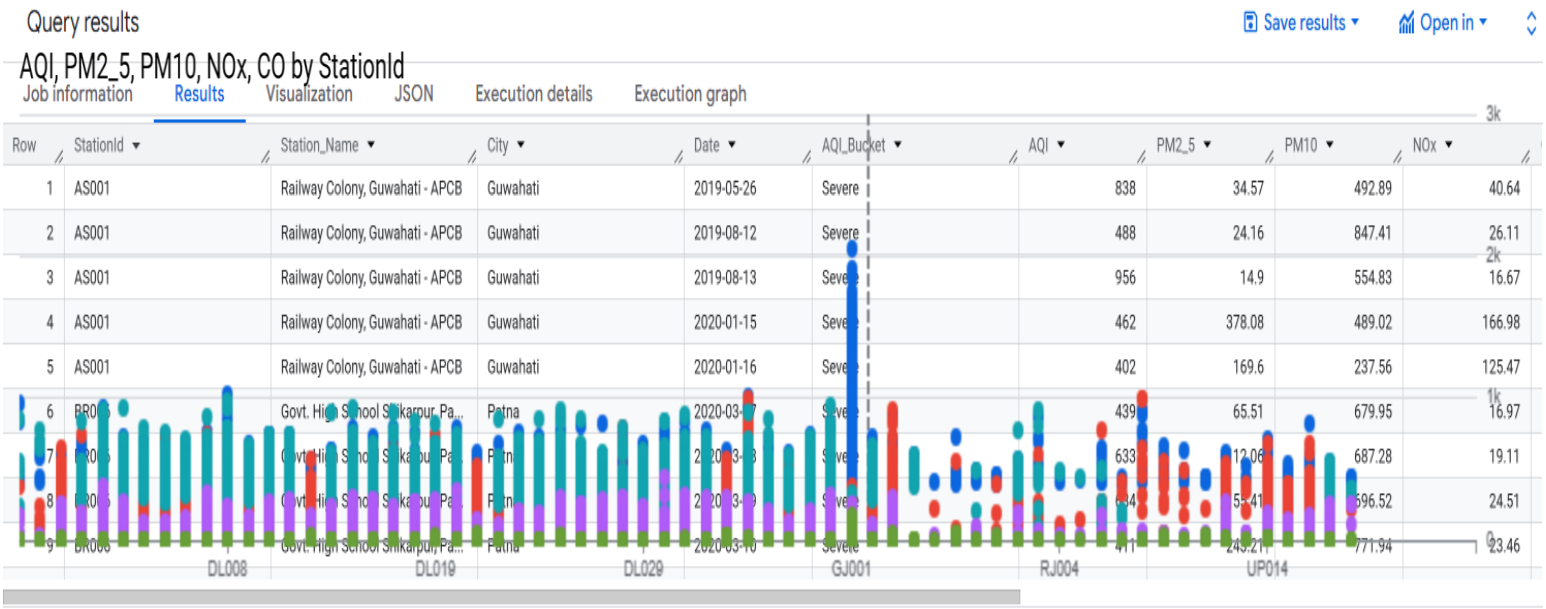
- Ahmedabad has the highest moderate to severe count, accounting for 5 instances or 31.25% of the total.
- Jorapokhar, Jaipur, Guwahati, and Patna each report the lowest count of moderate to severe cases, with 1 instance each, representing 6.25% individually and 25% collectively.
- The average moderate to severe count across all cities is 2.

## 13. List severe AQI events

Q. List all station-day records where AQI is above a threshold (e.g. “Severe” bucket). Include station, city, datetime,  pollutant , etc.

```
select StationId , Station_Name , City , Date , AQI_Bucket , AQI,
PM2_5 , PM10 , NOx ,CO ,O3 ,SO2
from `Air_quality_index.station_day` A
join `Air_quality_index.stations` B
on A.StationId = B.station_id
where AQI_Bucket = 'Severe'
```

## Result




## Visualization

## Insight

- AQI, PM2.5, and PM10 have decreased overall by 17.44%, 93.72%, and 19.9% respectively between 2015-1-1 and 2020-6-3.
- Significant decreases were observed in AQI, PM2.5 and PM10 around November 2019, dropping by 97.93%, 99.51%, and 99.26% respectively in a short period.
- Anand Vihar, Delhi - DPCC recorded the highest PM2.5 and PM10 values, accounting for 10.88% and 18.41% of the total, respectively.
- PM2.5 and PM10 increased from 0 in period (May 21, 2020 to May 27, 2020), to 26.72 and 486.99 in period (May 28, 2020 to Jun 3, 2020) respectively.

## 14. Minimum pollutant values

Q. For each pollutant (PM2_5, PM10, NO2, etc.), find the minimum non-null value recorded in 2020 year, and which station it occurred.

```
select StationId , Station_Name , City , Date ,
       min(AQI) as min_value_AQI, min(PM2_5) as min_value_PM2_5 ,
       min(PM10) as min_value_PM10 , min(NO2) as min_value_NO2 ,min(CO) as
min_value_CO
      ,min(O3) as min_value_O3 ,min(SO2) as min_value_SO2
 from `Air_quality_index.station_day` A
join `Air_quality_index.stations` B
```

```
on A.StationId = B.station_id
where Date between '2020-01-01' and '2020-12-31' and A.PM2_5 is not null
  and A.PM10 is not null
  and A.NO2 is not null
  and A.CO is not null
  and A.O3 is not null
  and A.SO2 is not null
group by StationId , Station_Name , City ,Date
```

## Result

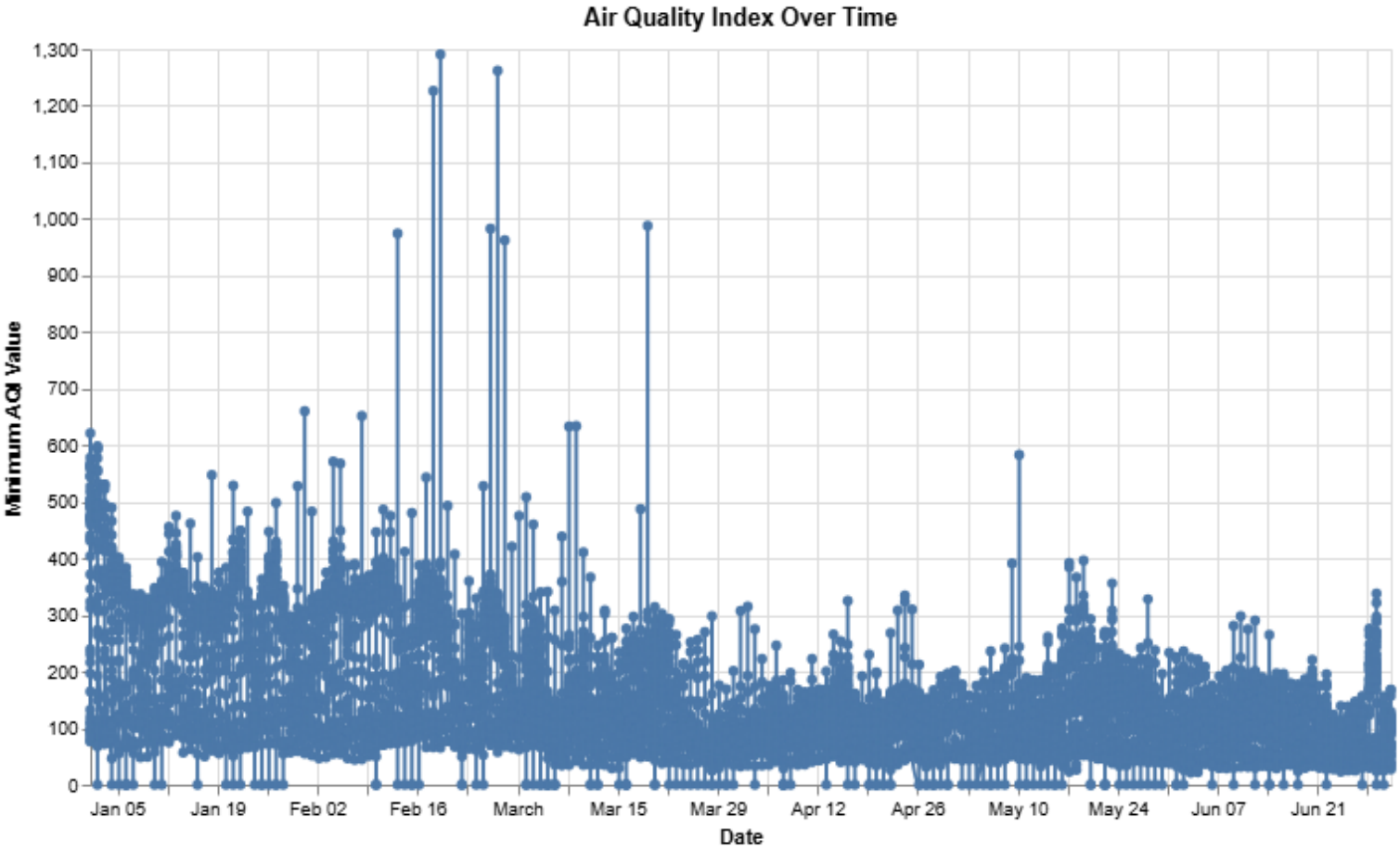
Query results

Save results | Open in

Job information | Results | Visualization | JSON | Execution details | Execution graph

| Row | StationId | Station_Name | City | Date | min_value_AQI | min_value_PM2_5 | min_value_PM10 | min_value_NO2 | min_value_CO | min_value_O3 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AP001 | Secretariat, Amaravati - APPCB | Amaravati | 2020-01-01 | 96 | 59.64 | 88.85 | 12.12 | 0.77 | |
| 2 | AP001 | Secretariat, Amaravati - APPCB | Amaravati | 2020-01-02 | 82 | 42.46 | 65.39 | 12.17 | 0.73 | |
| 3 | AP001 | Secretariat, Amaravati - APPCB | Amaravati | 2020-01-03 | 76 | 39.07 | 62.93 | 13.53 | 0.69 | |
| 4 | AP001 | Secretariat, Amaravati - APPCB | Amaravati | 2020-01-04 | 47 | 22.79 | 38.35 | 18.79 | 0.64 | |
| 5 | AP001 | Secretariat, Amaravati - APPCB | Amaravati | 2020-01-05 | 55 | 37.81 | 56.66 | 12.98 | 0.63 | |
| 6 | AP001 | Secretariat, Amaravati - APPCB | Amaravati | 2020-01-06 | 66 | 38.82 | 61.4 | 11.64 | 0.65 | |
| 7 | AP001 | Secretariat, Amaravati - APPCB | Amaravati | 2020-01-07 | 109 | 69.49 | 97.7 | 12.04 | 0.68 | |
| 8 | AP001 | Secretariat, Amaravati - APPCB | Amaravati | 2020-01-08 | 103 | 53.32 | 75.4 | 11.47 | 0.64 | |
| 9 | AP001 | Secretariat, Amaravati - APPCB | Amaravati | 2020-01-09 | 78 | 37.23 | 62.09 | 13.23 | 0.53 | |

## Visualization




## Insight

- Between January 1, 2020, and July 1, 2020, there was a significant decrease in overall minimum values for AQI (78.75%), PM2.5 (85.87%), and CO (70.36%).
- A strong positive correlation exists between min_value_AQI and min_value_PM2_5, indicating that as AQI increases, PM2.5 tends to increase as well.
- On July 1, 2020, Delhi recorded the highest minimum values for AQI, PM2.5 and CO, accounting for 71.74%, 77.61%, and 72.96% of the total, respectively.

## 15. Missing data detection

Q. For each station and month, count how many hours have NULL in any pollutant column.

```
select StationId , Station_Name ,format_datetime('%b' , Date) as month ,
sum(case when PM2_5 is null then 1 else 0 end) as null_count_of_PM2_5 ,
       sum(case when PM10 is null then 1 else 0 end ) as
null_count_of_PM10 ,
       sum(case when NO2 is null then 1 else 0 end ) as null_count_of_NO2,
       sum(case when CO is null then 1 else 0 end) as null_count_of_CO ,
       sum(case when SO2 is null then 1 else 0 end ) as null_count_of_SO2
from `Air_quality_index.station_day` A
join `Air_quality_index.stations` B
on A.StationId = B.station_id
group by StationId , Station_Name ,format_datetime('%b' ,date)
```

## Result

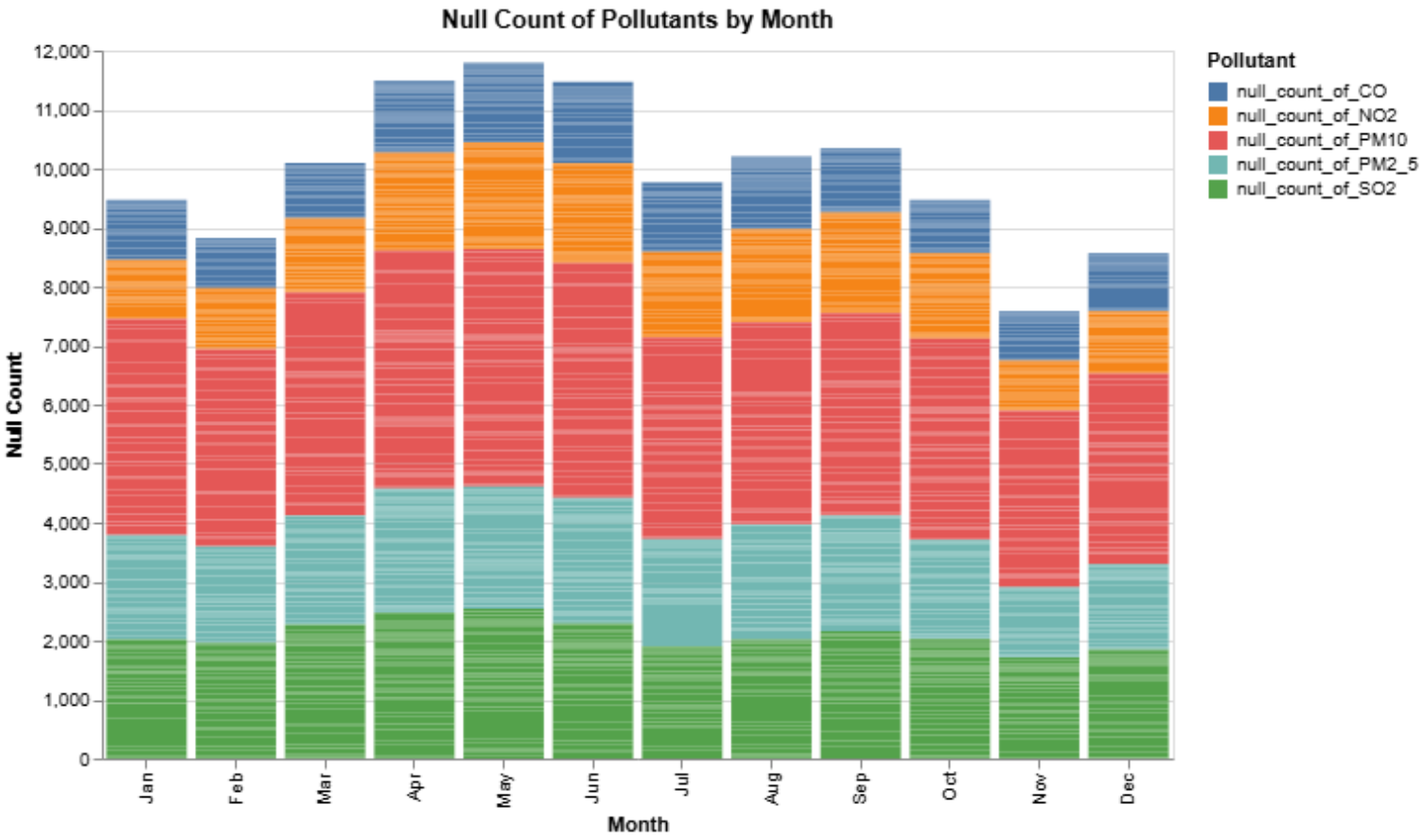
Query results

Save resul

Job information | Results | Visualization | JSON | Execution details | Execution graph

| Row | StationId | Station_Name | month | null_count_of_P... | null_count_of_P... | null_count_of_NO2 | null_count_of_CO | null_count_of_SO2 |
|---|---|---|---|---|---|---|---|---|
| 1 | AP001 | Secretariat, Amaravati - APPCB | Nov | 0 | 0 | 0 | 0 | 0 |
| 2 | AP001 | Secretariat, Amaravati - APPCB | Dec | 0 | 0 | 0 | 0 | 0 |
| 3 | AP001 | Secretariat, Amaravati - APPCB | Jan | 0 | 0 | 0 | 0 | 0 |
| 4 | AP001 | Secretariat, Amaravati - APPCB | Feb | 0 | 0 | 0 | 2 | 0 |
| 5 | AP001 | Secretariat, Amaravati - APPCB | Mar | 0 | 0 | 0 | 0 | 0 |
| 6 | AP001 | Secretariat, Amaravati - APPCB | Apr | 0 | 0 | 0 | 14 | 9 |
| 7 | AP001 | Secretariat, Amaravati - APPCB | May | 5 | 2 | 2 | 28 | 3 |
| 8 | AP001 | Secretariat, Amaravati - APPCB | Jun | 0 | 0 | 0 | 0 | 0 |
| 9 | AP001 | Secretariat, Amaravati - APPCB | Jul | 0 | 0 | 0 | 0 | 0 |

## Visualization




## Insight

- However, there were significant increases in null counts during specific periods, such as CO increasing by 61.05% from February to June and PM10 increasing by 20.75% from February to April.
- East Arjun Nagar, Delhi - CPCB station recorded the highest null_count_of_CO, accounting for 11.95% of the total, while Sanegurava Halli, Bengaluru - KSPCB station recorded the highest null_count_of_PM2_5, accounting for 7.86% of the total.
- A positive correlation exists between null counts of PM2.5 and PM10, and between CO and PM2.5, suggesting a potential relationship in data availability across these pollutants.

---------- END ----------