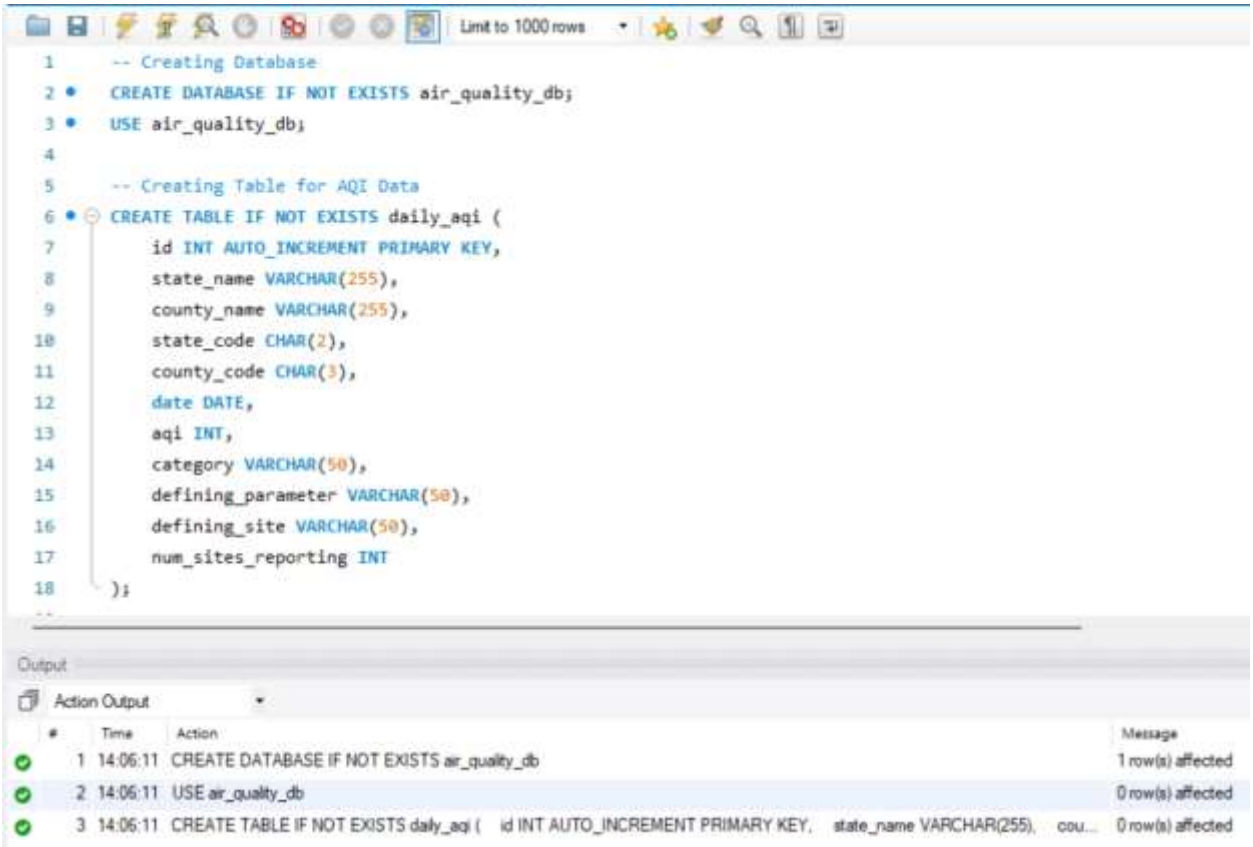
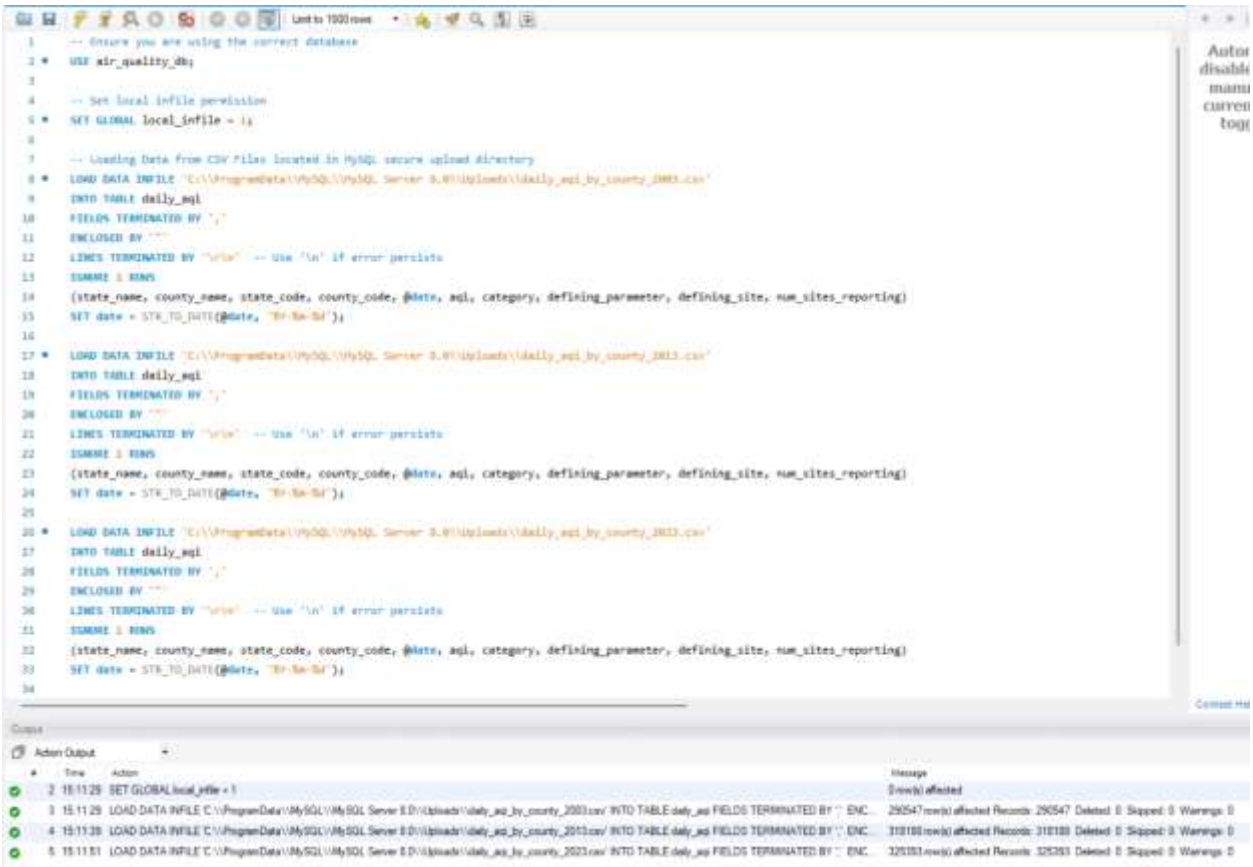
Name : Meenakshi H
Submission Date: 3/16/2025
UNID: U1526460
Course Name : Data Base Theory & Design
Submission Title : Air Quality Data

```sql
-- Creating Database
CREATE DATABASE IF NOT EXISTS air_quality_db;
USE air_quality_db;

-- Creating Table for AQI Data
CREATE TABLE IF NOT EXISTS daily_aqi (
    id INT AUTO_INCREMENT PRIMARY KEY,
    state_name VARCHAR(255),
    county_name VARCHAR(255),
    state_code CHAR(2),
    county_code CHAR(3),
    date DATE,
    aqi INT,
    category VARCHAR(50),
    defining_parameter VARCHAR(50),
    defining_site VARCHAR(50),
    num_sites_reporting INT
);
```

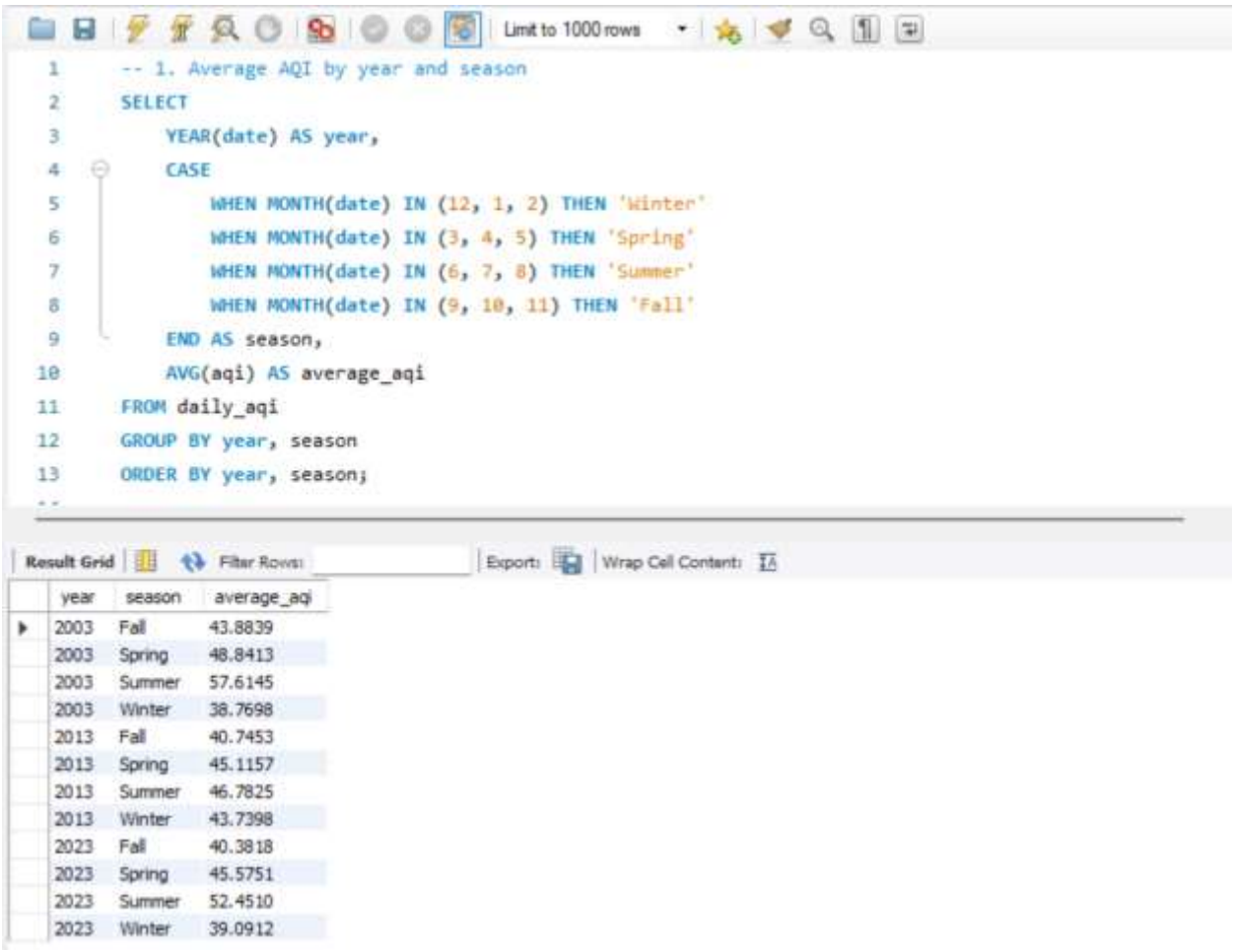



```sql
-- Ensure you are using the correct database
USE air_quality_db;

-- Set local infile permission
SET GLOBAL local_infile = 1;
```

```sql
-- Loading Data from CSV Files located in MySQL secure upload directory

LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server
8.0\\Uploads\\daily_aqi_by_county_2003.csv'
INTO TABLE daily_aqi
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'  -- Use '\n' if error persists
IGNORE 1 ROWS
(state_name, county_name, state_code, county_code, @date, aqi, category,
defining_parameter, defining_site, num_sites_reporting)
SET date = STR_TO_DATE(@date, '%Y-%m-%d');

LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server
8.0\\Uploads\\daily_aqi_by_county_2013.csv'
INTO TABLE daily_aqi
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'  -- Use '\n' if error persists
IGNORE 1 ROWS
(state_name, county_name, state_code, county_code, @date, aqi, category,
defining_parameter, defining_site, num_sites_reporting)
SET date = STR_TO_DATE(@date, '%Y-%m-%d');

LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server
8.0\\Uploads\\daily_aqi_by_county_2023.csv'
INTO TABLE daily_aqi
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'  -- Use '\n' if error persists
IGNORE 1 ROWS
(state_name, county_name, state_code, county_code, @date, aqi, category,
defining_parameter, defining_site, num_sites_reporting)
SET date = STR_TO_DATE(@date, '%Y-%m-%d');
```



```
-- Queries to Extract Insights

-- 1. Average AQI by year and season
SELECT
    YEAR(date) AS year,
    CASE
        WHEN MONTH(date) IN (12, 1, 2) THEN 'Winter'
        WHEN MONTH(date) IN (3, 4, 5) THEN 'Spring'
        WHEN MONTH(date) IN (6, 7, 8) THEN 'Summer'
        WHEN MONTH(date) IN (9, 10, 11) THEN 'Fall'
    END AS season,
    AVG(aqi) AS average_aqi
FROM daily_aqi
GROUP BY year, season
ORDER BY year, season;
```

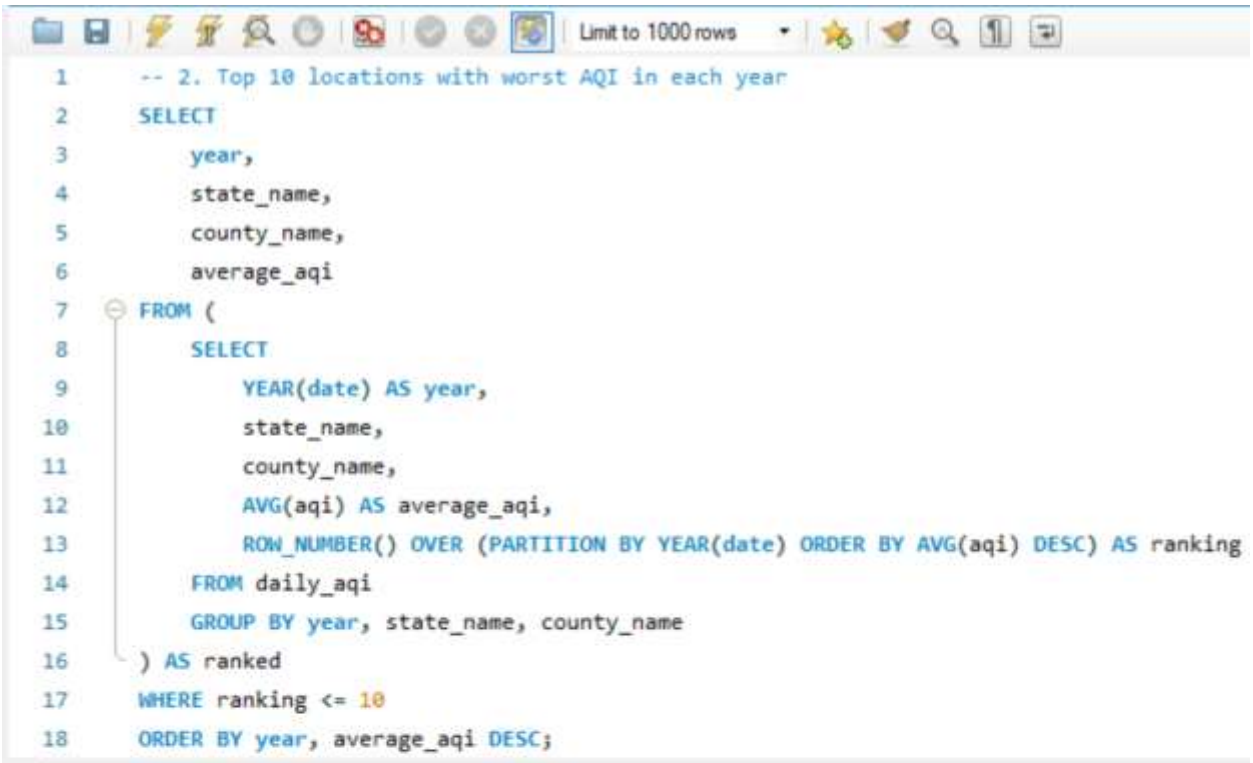


```
1   -- 1. Average AQI by year and season
2   SELECT
3       YEAR(date) AS year,
4       CASE
5           WHEN MONTH(date) IN (12, 1, 2) THEN 'Winter'
6           WHEN MONTH(date) IN (3, 4, 5) THEN 'Spring'
7           WHEN MONTH(date) IN (6, 7, 8) THEN 'Summer'
8           WHEN MONTH(date) IN (9, 10, 11) THEN 'Fall'
9       END AS season,
10      AVG(aqi) AS average_aqi
11  FROM daily_aqi
12  GROUP BY year, season
13  ORDER BY year, season;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

year	season	average_aqi
2003	Fall	43.8839
2003	Spring	48.8413
2003	Summer	57.6145
2003	Winter	38.7698
2013	Fall	40.7453
2013	Spring	45.1157
2013	Summer	46.7825
2013	Winter	43.7398
2023	Fall	40.3818
2023	Spring	45.5751
2023	Summer	52.4510
2023	Winter	39.0912

```sql
-- 2. Top 10 locations with worst AQI in each year
SELECT
    year,
    state_name,
    county_name,
    average_aqi
FROM (
    SELECT
        YEAR(date) AS year,
        state_name,
        county_name,
        AVG(aqi) AS average_aqi,
        ROW_NUMBER() OVER (PARTITION BY YEAR(date) ORDER BY AVG(aqi) DESC) AS ranking
    FROM daily_aqi
    GROUP BY year, state_name, county_name
) AS ranked
WHERE ranking <= 10
ORDER BY year, average_aqi DESC;
```



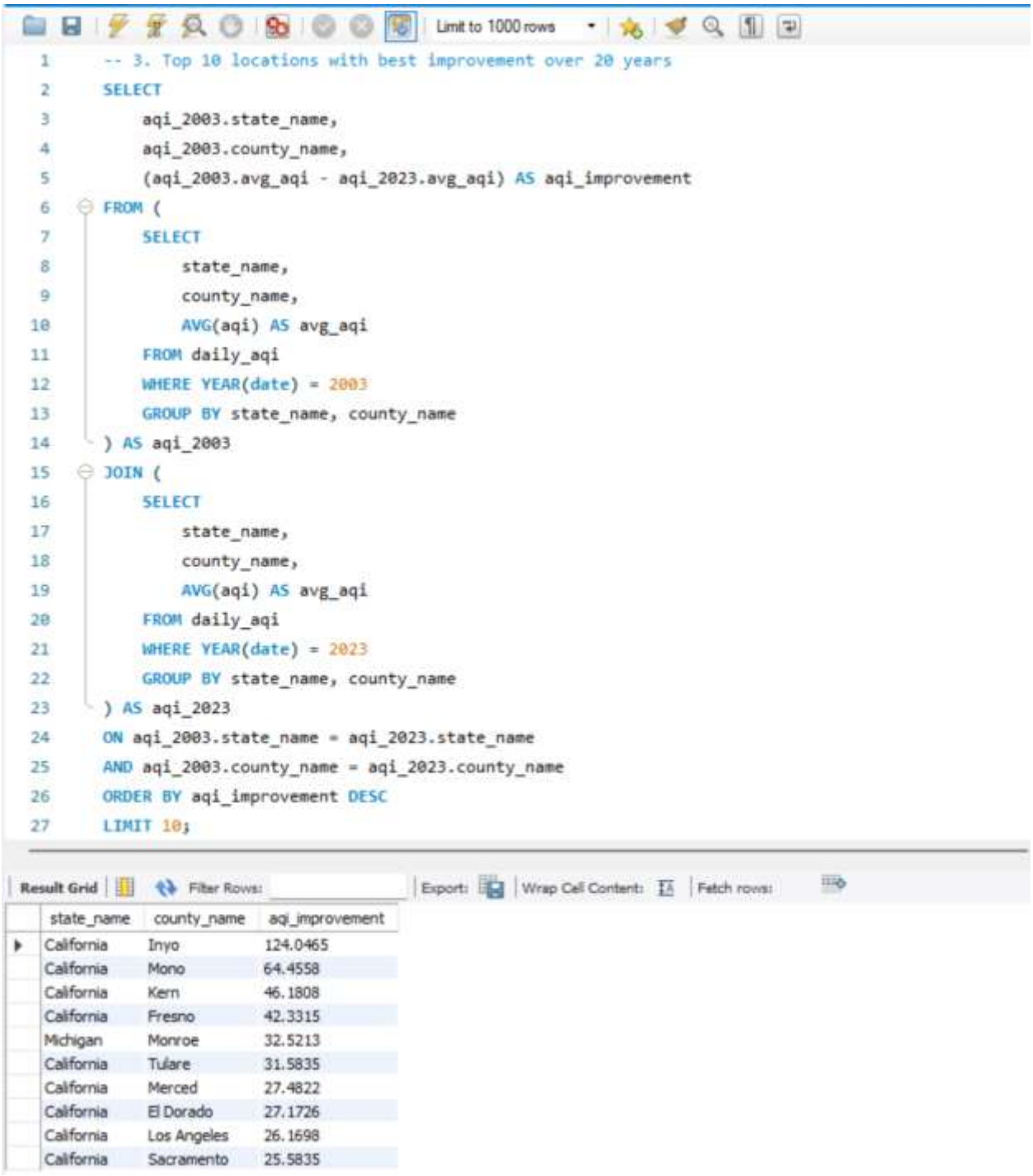
```sql
-- 2. Top 10 locations with worst AQI in each year
SELECT
    year,
    state_name,
    county_name,
    average_aqi
FROM (
    SELECT
        YEAR(date) AS year,
        state_name,
        county_name,
        AVG(aqi) AS average_aqi,
        ROW_NUMBER() OVER (PARTITION BY YEAR(date) ORDER BY AVG(aqi) DESC) AS ranking
    FROM daily_aqi
    GROUP BY year, state_name, county_name
) AS ranked
WHERE ranking <= 10
ORDER BY year, average_aqi DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

year	state_name	county_name	average_aqi
2003	California	Inyo	196.7479
2003	California	Kern	122.3644
2003	California	Fresno	113.2712
2003	California	Riverside	110.8575
2003	California	San Bernardino	110.1288
2003	California	Los Angeles	109.3890
2003	California	Tulare	108.2630
2003	California	Mono	103.4777
2003	Country Of Mexico	BAJA CALIFORNIA NORTE	100.8384
2003	California	Merced	83.9808
2013	California	Riverside	96.0658
2013	California	Kern	94.8932
2013	California	San Bernardino	94.3507
2013	Nevada	Washoe	93.6795
2013	California	Fresno	93.6027
2013	California	Tulare	91.7151
2013	California	Los Angeles	89.2055
2013	California	Kings	82.8822
2013	Arizona	Pinal	80.9370
2013	California	Madera	79.7397

```sql
-- 3. Top 10 locations with best improvement over 20 years
SELECT
    aqi_2003.state_name,
    aqi_2003.county_name,
```

```sql
    (aqi_2003.avg_aqi - aqi_2023.avg_aqi) AS aqi_improvement
FROM (
    SELECT
        state_name,
        county_name,
        AVG(aqi) AS avg_aqi
    FROM daily_aqi
    WHERE YEAR(date) = 2003
    GROUP BY state_name, county_name
) AS aqi_2003
JOIN (
    SELECT
        state_name,
        county_name,
        AVG(aqi) AS avg_aqi
    FROM daily_aqi
    WHERE YEAR(date) = 2023
    GROUP BY state_name, county_name
) AS aqi_2023
ON aqi_2003.state_name = aqi_2023.state_name
AND aqi_2003.county_name = aqi_2023.county_name
ORDER BY aqi_improvement DESC
LIMIT 10;
```



```
-- 3. Top 10 locations with best improvement over 20 years
SELECT
    aqi_2003.state_name,
    aqi_2003.county_name,
    (aqi_2003.avg_aqi - aqi_2023.avg_aqi) AS aqi_improvement
FROM (
    SELECT
        state_name,
        county_name,
        AVG(aqi) AS avg_aqi
    FROM daily_aqi
    WHERE YEAR(date) = 2003
    GROUP BY state_name, county_name
) AS aqi_2003
JOIN (
    SELECT
        state_name,
        county_name,
        AVG(aqi) AS avg_aqi
    FROM daily_aqi
    WHERE YEAR(date) = 2023
    GROUP BY state_name, county_name
) AS aqi_2023
ON aqi_2003.state_name = aqi_2023.state_name
AND aqi_2003.county_name = aqi_2023.county_name
ORDER BY aqi_improvement DESC
LIMIT 10;
```

Result Grid

state_name	county_name	aqi_improvement
California	Inyo	124.0465
California	Mono	64.4558
California	Kern	46.1808
California	Fresno	42.3315
Michigan	Monroe	32.5213
California	Tulare	31.5835
California	Merced	27.4822
California	El Dorado	27.1726
California	Los Angeles	26.1698
California	Sacramento	25.5835

```
-- 4. Top 10 locations with worst decline over 20 years
SELECT
    aqi_2003.state_name,
    aqi_2003.county_name,
    (aqi_2023.avg_aqi - aqi_2003.avg_aqi) AS aqi_decline
FROM (
    SELECT
```

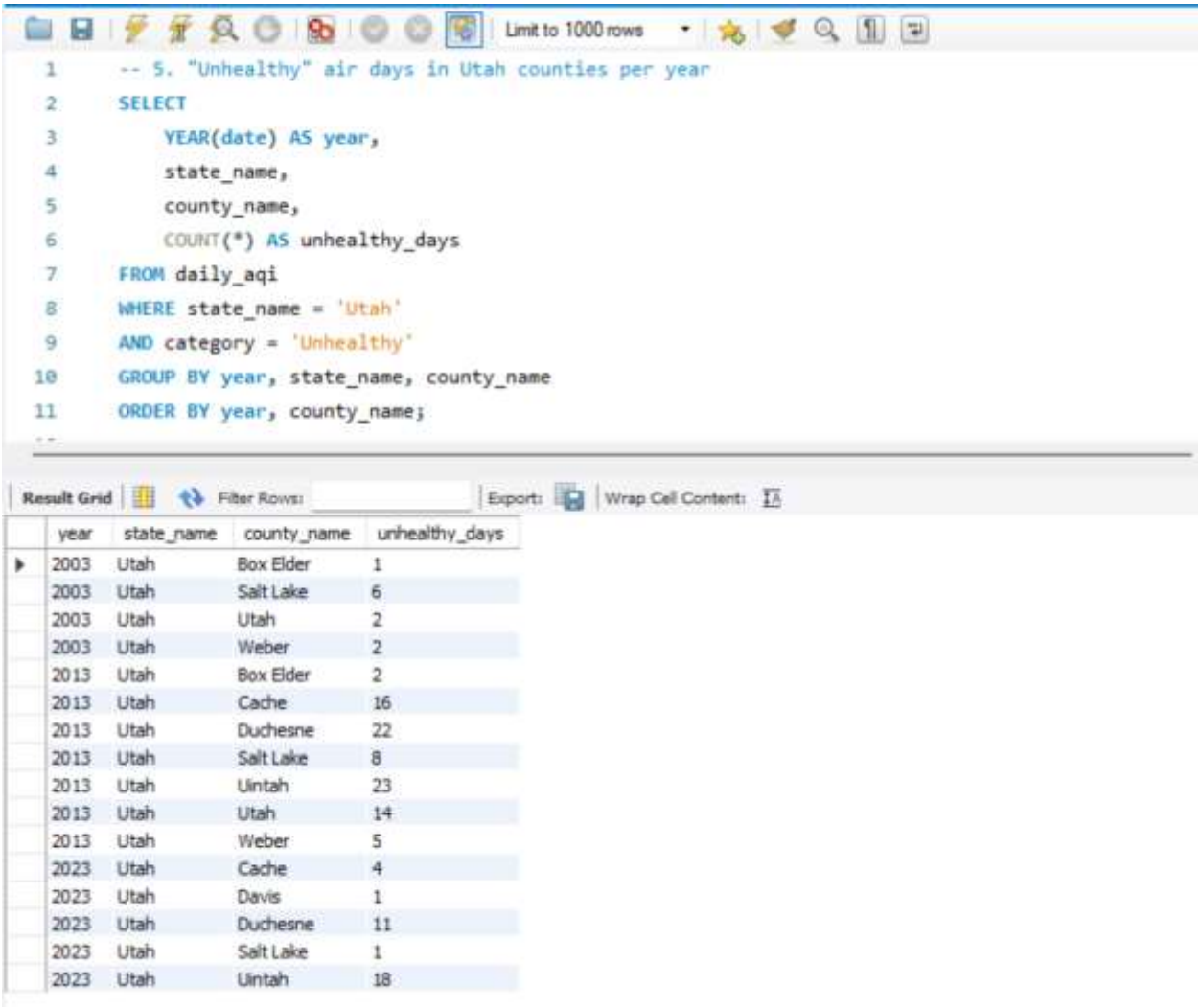
```sql
        state_name,
        county_name,
        AVG(aqi) AS avg_aqi
    FROM daily_aqi
    WHERE YEAR(date) = 2003
    GROUP BY state_name, county_name
) AS aqi_2003
JOIN (
    SELECT
        state_name,
        county_name,
        AVG(aqi) AS avg_aqi
    FROM daily_aqi
    WHERE YEAR(date) = 2023
    GROUP BY state_name, county_name
) AS aqi_2023
ON aqi_2003.state_name = aqi_2023.state_name
AND aqi_2003.county_name = aqi_2023.county_name
ORDER BY aqi_decline DESC
LIMIT 10;
```

```sql
-- 4. Top 10 locations with worst decline over 20 years
SELECT
    aqi_2003.state_name,
    aqi_2003.county_name,
    (aqi_2023.avg_aqi - aqi_2003.avg_aqi) AS aqi_decline
FROM (
    SELECT
        state_name,
        county_name,
        AVG(aqi) AS avg_aqi
    FROM daily_aqi
    WHERE YEAR(date) = 2003
    GROUP BY state_name, county_name
) AS aqi_2003
JOIN (
    SELECT
        state_name,
        county_name,
        AVG(aqi) AS avg_aqi
    FROM daily_aqi
    WHERE YEAR(date) = 2023
    GROUP BY state_name, county_name
) AS aqi_2023
ON aqi_2003.state_name = aqi_2023.state_name
AND aqi_2003.county_name = aqi_2023.county_name
ORDER BY aqi_decline DESC
LIMIT 10;
```

state_name	county_name	aqi_decline
Idaho	Bannock	55.9594
Utah	Uintah	55.0448
Utah	Duchesne	52.1935
Colorado	Garfield	40.7433
Wisconsin	Forest	35.1885
Wisconsin	Eau Claire	34.0629
Colorado	Jackson	33.5725
North Dakota	Burke	33.1598
Montana	Cascade	33.0260
Kansas	Neosho	31.2948

```sql
-- 5. "Unhealthy" air days in Utah counties per year
SELECT
    YEAR(date) AS year,
    state_name,
    county_name,
    COUNT(*) AS unhealthy_days
FROM daily_aqi
WHERE state_name = 'Utah'
AND category = 'Unhealthy'
```

```
GROUP BY year, state_name, county_name
ORDER BY year, county_name;
```

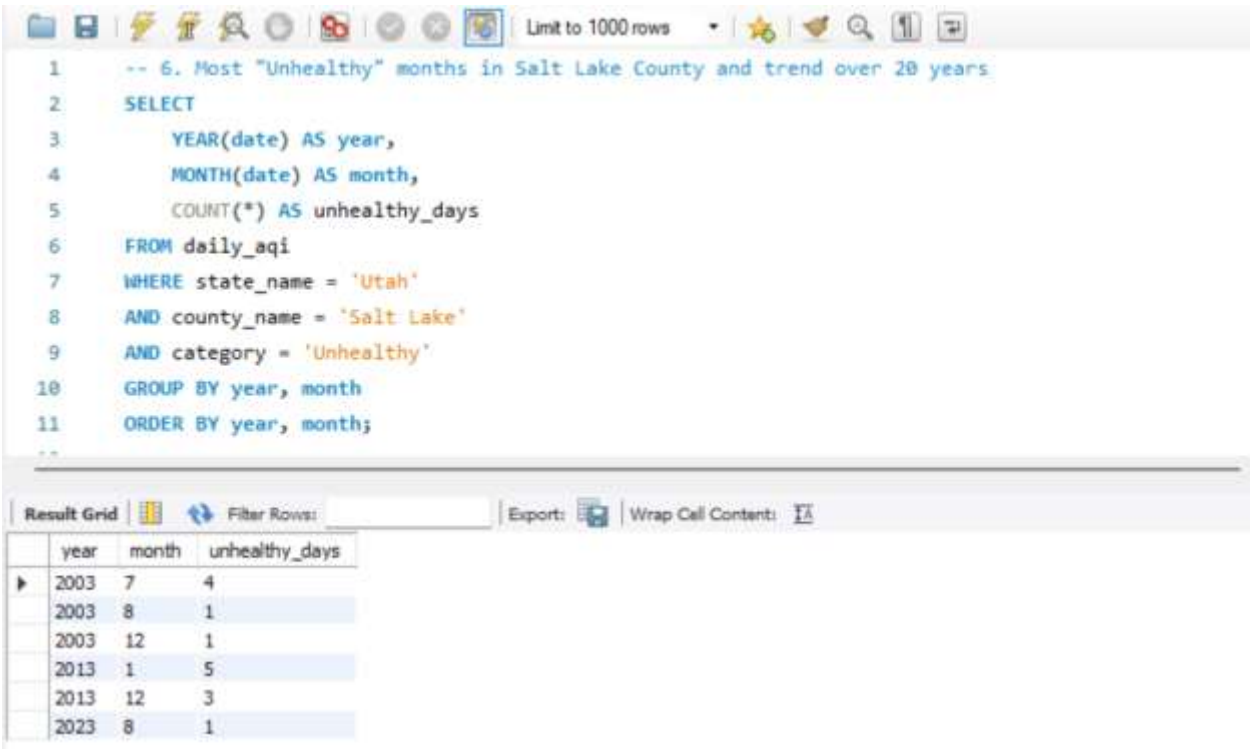


```
-- 5. "Unhealthy" air days in Utah counties per year
SELECT
    YEAR(date) AS year,
    state_name,
    county_name,
    COUNT(*) AS unhealthy_days
FROM daily_aqi
WHERE state_name = 'Utah'
AND category = 'Unhealthy'
GROUP BY year, state_name, county_name
ORDER BY year, county_name;
```

Result Grid

year	state_name	county_name	unhealthy_days
2003	Utah	Box Elder	1
2003	Utah	Salt Lake	6
2003	Utah	Utah	2
2003	Utah	Weber	2
2013	Utah	Box Elder	2
2013	Utah	Cache	16
2013	Utah	Duchesne	22
2013	Utah	Salt Lake	8
2013	Utah	Uintah	23
2013	Utah	Utah	14
2013	Utah	Weber	5
2023	Utah	Cache	4
2023	Utah	Davis	1
2023	Utah	Duchesne	11
2023	Utah	Salt Lake	1
2023	Utah	Uintah	18

```
-- 6. Most "Unhealthy" months in Salt Lake County and trend over 20 years
SELECT
    YEAR(date) AS year,
    MONTH(date) AS month,
    COUNT(*) AS unhealthy_days
FROM daily_aqi
WHERE state_name = 'Utah'
AND county_name = 'Salt Lake'
AND category = 'Unhealthy'
GROUP BY year, month
ORDER BY year, month;
```



```sql
-- 6. Most "Unhealthy" months in Salt Lake County and trend over 20 years
SELECT
    YEAR(date) AS year,
    MONTH(date) AS month,
    COUNT(*) AS unhealthy_days
FROM daily_aqi
WHERE state_name = 'Utah'
AND county_name = 'Salt Lake'
AND category = 'Unhealthy'
GROUP BY year, month
ORDER BY year, month;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

year	month	unhealthy_days
2003	7	4
2003	8	1
2003	12	1
2013	1	5
2013	12	3
2023	8	1

-

Executive Summary: Air Quality Trends in Utah (2003-2023)

Governor and Esteemed Leaders,

The analysis of air quality data from 2003 to 2023 highlights critical trends in seasonal pollution levels, worst-affected locations, areas with significant improvement or decline, and the trend of unhealthy air days in Utah counties. The findings underscore the need for targeted interventions to mitigate worsening air quality in specific regions.

Key Findings

1. **Seasonal Air Quality Variations**
    - AQI levels are highest in **summer** and lowest in **winter**, reflecting increased ozone formation and wildfire activity.
    - The **worst summer AQI** was recorded in 2003, averaging **57.61**, while the **lowest winter AQI** was in 2023 at **39.09**.
2. **Worst-Affected Counties**
    - California counties (Inyo, Kern, Riverside, and Fresno) dominate the list of highest AQI values.
    - Utah's **Duchesne, Uintah, and Salt Lake counties** have seen periods of poor air quality, warranting attention.
3. **Improvement vs. Decline Over 20 Years**
    - **Most Improved:** Inyo County, CA (**124-point AQI reduction**), followed by Mono and Kern.
    - **Biggest Decline:** Bannock County, ID, and **Duchesne and Uintah in Utah**, with AQI worsening by over **50 points** since 2003.
4. **Unhealthy Air Days in Utah**
    - **Duchesne and Uintah counties** saw unhealthy air days increase significantly from **2003 to 2013**, largely due to oil and gas emissions.
    - **Salt Lake County improved in 2023**, showing a decline in unhealthy air days compared to past decades.
5. **Salt Lake County's Long-Term Trends**
    - **July and December were historically the worst months** for air pollution.
    - **2023 data indicates improvement**, with fewer unhealthy days recorded compared to previous years.

Policy Recommendations

- **Targeted Emission Controls:** Implement stricter **industrial and vehicular emission policies** in **Duchesne, Uintah, and Salt Lake counties**, where air quality has worsened.
- **Wildfire & Ozone Mitigation:** Expand **forest management and wildfire prevention** strategies to control seasonal spikes in AQI.
- **Clean Energy Incentives:** Encourage **renewable energy adoption** and **transition away from fossil fuel emissions** in high-risk areas.
- **Stronger Public Awareness:** Promote carpooling, remote work, and public transit during peak pollution months.
- **Enhanced Monitoring & Regulations:** Strengthen air quality monitoring, particularly in **northern and eastern Utah**, where air quality trends show increasing risks.

Utah can continue to improve its air quality and protect public health by implementing these measures. The data suggests progress in some areas but also highlights urgent concerns requiring immediate policy action.