Varun Selvam

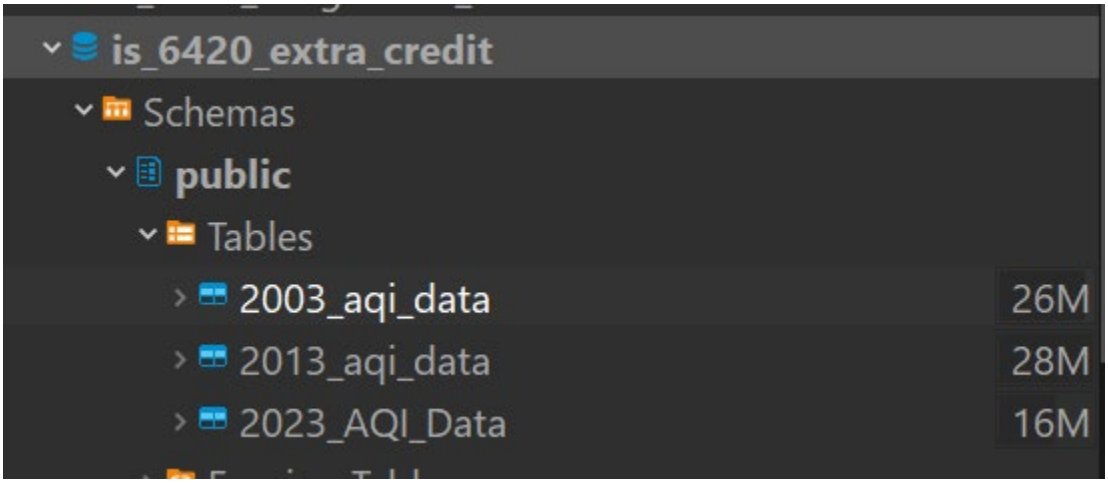
March 17th, 2024

U0676196

IS 6420 – Database Theory and Design

Air Quality Data Analysis

STEPS 1, 2, AND 3:



QUESTION 1: What is the average AQI (air quality index) by year by season (winter, spring, summer, fall)?

```
select extract(year from recorded_date) as
recorded_year,
case
    when extract(month from recorded_date) in
(12,1,2,3) then 'WINTER'
    when extract(month from recorded_date) in
(4,5) then 'SPRING'
    when extract(month from recorded_date) in
(6,7,8) then 'SUMMER'
    when extract(month from recorded_date) in
(9,10,11) then 'FALL'
end as SEASON,
AVG(AQI)
from "2003_aqi_data"
group by recorded_year, SEASON
```

select extract(year from recorded_d

	recorded_year	season	avg
1	2,003	FALL	41.7328973565
2	2,003	SPRING	49.2912015453
3	2,003	SUMMER	56.0940187623
4	2,003	WINTER	37.3841122289

```
select extract(year from recorded_date) as
recorded_year,
case
    when extract(month from recorded_date) in
(12,1,2,3) then 'WINTER'
    when extract(month from recorded_date) in
(4,5) then 'SPRING'
    when extract(month from recorded_date) in
(6,7,8) then 'SUMMER'
    when extract(month from recorded_date) in
(9,10,11) then 'FALL'
end as SEASON,
AVG(AQI)
from "2013_aqi_data"
group by recorded_year, SEASON
```

	recorded_year	season	avg
1	2,013	FALL	37.5120609767
2	2,013	SPRING	43.614516101
3	2,013	SUMMER	44.0148792003
4	2,013	WINTER	39.7642449464

```
select extract(year from recorded_date) as
recorded_year,
case
    when extract(month from recorded_date) in
(12,1,2,3) then 'WINTER'
    when extract(month from recorded_date) in
(4,5) then 'SPRING'
    when extract(month from recorded_date) in
(6,7,8) then 'SUMMER'
    when extract(month from recorded_date) in
(9,10,11) then 'FALL'
end as SEASON,
AVG(AQI)
from "2023_AQI_Data"
group by recorded_year, SEASON
```

	recorded_year	season	avg
1	2,023	FALL	38.1925817267
2	2,023	SPRING	45.2395008139
3	2,023	SUMMER	51.3675645233
4	2,023	WINTER	37.0167705753

- **QUESTION 2: What were the top 10 locations with worst AQI in each year?**

```
select recorded_year, state_name,
county_name, category, aqi from
```

```
(select extract(year from recorded_date) as
recorded_year, state_name, county_name,
category, aqi
from "2013_aqi_data"
order by aqi desc) worst_aqi
limit 10
```

	recorded_year	state_name	county_name	category	aqi
1	2,013	California	Mono	Hazardous	3,180
2	2,013	California	Mono	Hazardous	1,766
3	2,013	Colorado	Prowers	Hazardous	1,116
4	2,013	Arizona	Pinal	Hazardous	902
5	2,013	Nevada	Washoe	Hazardous	895
6	2,013	Nevada	Washoe	Hazardous	895
7	2,013	Nevada	Washoe	Hazardous	895
8	2,013	Nevada	Washoe	Hazardous	895
9	2,013	Nevada	Washoe	Hazardous	895
10	2,013	Nevada	Washoe	Hazardous	895

```
select recorded_year, state_name, county_name,
category, aqi from
(select extract(year from recorded_date) as
recorded_year, state_name, county_name,
category, aqi
from "2003_aqi_data"
order by aqi desc) worst_aqi
limit 10
```

select recorded_year, state_name, c | Enter a SQL expression to filter results (

	recorded_year	state_name	county_name	category	aqi
1	2,003	California	Inyo	Hazardous	16,515
2	2,003	California	Inyo	Hazardous	6,488
3	2,003	California	Mono	Hazardous	5,641
4	2,003	California	Mono	Hazardous	5,179
5	2,003	California	Inyo	Hazardous	3,481
6	2,003	California	Inyo	Hazardous	2,416
7	2,003	Country Of Mexico	CHIHUAHUA STATE	Hazardous	2,405
8	2,003	California	Inyo	Hazardous	2,296
9	2,003	California	Inyo	Hazardous	2,223
10	2,003	California	Inyo	Hazardous	2,161

```
select recorded_year, state_name, county_name,
category, aqi from
(select extract(year from recorded_date) as
recorded_year, state_name, county_name,
category, aqi
from "2023_AQI_Data"
order by aqi desc) worst_aqi
limit 10
```

	recorded_year	state_name	county_name	category	aqi
1	2,023	California	Mono	Hazardous	1,695
2	2,023	California	Mono	Hazardous	1,391
3	2,023	California	Mono	Hazardous	820
4	2,023	California	Inyo	Hazardous	757
5	2,023	California	Riverside	Hazardous	644
6	2,023	California	Inyo	Hazardous	597
7	2,023	California	Inyo	Hazardous	596
8	2,023	Idaho	Bannock	Hazardous	551
9	2,023	Idaho	Bannock	Hazardous	551
10	2,023	Idaho	Bannock	Hazardous	551

- **Question 3:** What were the top 10 locations that had **the best improvement** over 20 years, from the first year to the most recent year?

```
select state_name, county_name, "2023_aqi",
"2003_aqi",diff_in_aqi from
(select state_name, county_name,
sum(max_aqi) "2023_aqi", sum(min_aqi)
"2003_aqi",
(sum(max_aqi) - sum(min_aqi)) as
diff_in_aqi
from
(
select extract(year from recorded_date) as
rec_year,
state_name, county_name, max(aqi) max_aqi,
0 as  min_aqi
from "2023_AQI_Data"
--where county_name = 'Fresno'
group by rec_year, state_name, county_name,
min_aqi
union
select extract(year from recorded_date)
rec_year, state_name, county_name, 0 as
max_aqi, min(aqi) min_aqi
from "2003_aqi_data"
--where county_name = 'Fresno'
group by rec_year, state_name, county_name,
max_aqi
) combined_aqi
group by state_name, county_name
```

```
order by diff_in_aqi asc
) final_aqi
limit 10
```

	state_name	county_name	2023_aqi	2003_aqi	diff_in_aqi
1	Country Of Mexico	BAJA CALIFORNIA NORTE	0	32	-32
2	Wisconsin	Green	0	24	-24
3	Oklahoma	Payne	0	23	-23
4	Maine	MOBILE MONITORS	0	22	-22
5	California	Modoc	0	21	-21
6	Kentucky	Laurel	0	21	-21
7	North Carolina	Camden	0	21	-21
8	Michigan	Leelanau	0	20	-20
9	Mississippi	Jones	0	20	-20
10	Alabama	Houston	0	20	-20

- Question 3: What were the **10 locations with the worst decline over 20 years**?

```
select state_name, county_name, "2023_aqi",
"2003_aqi",diff_in_aqi from
(select state_name, county_name,
sum(max_aqi) "2023_aqi", sum(min_aqi)
"2003_aqi",
(sum(max_aqi) - sum(min_aqi)) as
diff_in_aqi
from
(
select extract(year from recorded_date) as
rec_year,
state_name, county_name, max(aqi) max_aqi,
0 as  min_aqi
from "2023_AQI_Data"
--where county_name = 'Fresno'
```

```
group by rec_year, state_name, county_name,
min_aqi
union
select extract(year from recorded_date)
rec_year, state_name, county_name, 0 as
max_aqi, min(aqi) min_aqi
from "2003_aqi_data"
--where county_name = 'Fresno'
group by rec_year, state_name, county_name,
max_aqi
) combined_aqi
group by state_name, county_name
order by diff_in_aqi desc
) final_aqi
limit 10
```

select state_name, county_name, "2 | Enter a SQL expression to filter results (use Ctrl+S

	state_name	county_name	2023_aqi	2003_aqi	diff_in_aqi
1	California	Mono	1,695	2	1,693
2	California	Inyo	757	12	745
3	California	Riverside	644	32	612
4	Idaho	Bannock	551	3	548
5	Arizona	Yuma	517	8	509
6	California	Imperial	505	25	480
7	Oregon	Deschutes	332	5	327
8	Pennsylvania	Northampton	309	15	294
9	Pennsylvania	Lehigh	293	6	287
10	North Dakota	Oliver	286	1	285

Question 4: In Utah counties, how many days of "Unhealthy" air did we have in each year? Is it improving?

```
select extract(year from recorded_date),
state_name, county_name, count(category)
```

```
from "2003_aqi_data"
where state_name = 'Utah'
and county_name = 'Utah'
and category like 'Unhealthy%'
group by extract(year from recorded_date),
state_name, county_name
```

select extract(year from recorded_d | Enter a SQL expression to filter res

	extract	state_name	county_name	count
1	2,003	Utah	Utah	30

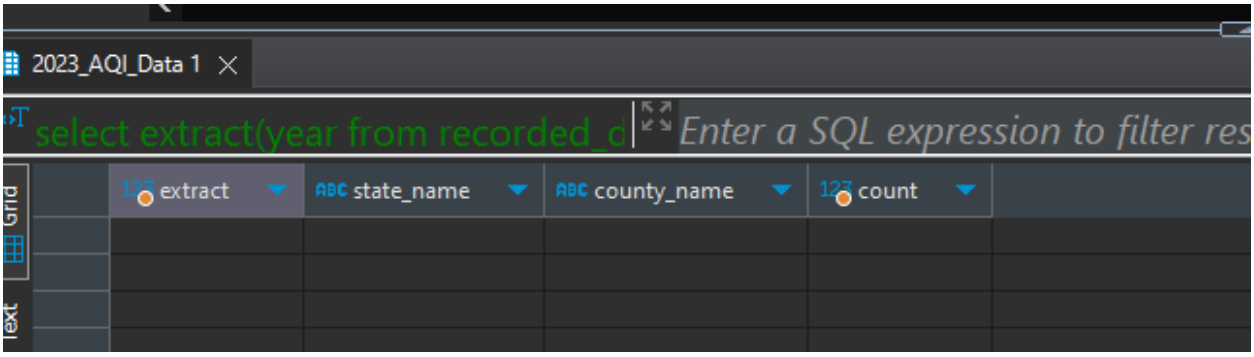
```
select extract(year from recorded_date),
state_name, county_name, count(category)
from "2013_aqi_data"
where state_name = 'Utah'
and county_name = 'Utah'
and category like 'Unhealthy%'
group by extract(year from recorded_date),
state_name, county_name
```

select extract(year from recorded_d | Enter a SQL expression to

	extract	state_name	county_name	count
1	2,013	Utah	Utah	41

```
select extract(year from recorded_date),
state_name, county_name, count(category)
from "2023_AQI_Data"
```

```
where state_name = 'Utah'
and county_name = 'Utah'
and category like 'Unhealthy%'
group by extract(year from recorded_date),
state_name, county_name
```



2023 did not have any unhealthy days, 2003 had 30 unhealthy days. But in 2013, there were 41 unhealthy days. Seems like it is improving.

Question 5: In Salt Lake County, which months have the most "Unhealthy" days? Has that changed in 20 years?

2003

```
select extract(year from recorded_date)
rec_year,
extract(month from recorded_date) rec_month,
state_name, county_name, count(category)
from "2003_aqi_data"
where state_name = 'Utah'
and county_name = 'Salt Lake'
and category like 'Unhealthy%'
group by extract(year from recorded_date),
```

```
extract(month from recorded_date), state_name,
county_name
```

select extract(year from recorded_d | Enter a SQL expression to filter results

	rec_year	rec_month	state_name	county_name	count
1	2,003	1	Utah	Salt Lake	2
2	2,003	2	Utah	Salt Lake	1
3	2,003	4	Utah	Salt Lake	1
4	2,003	5	Utah	Salt Lake	2
5	2,003	6	Utah	Salt Lake	6
6	2,003	7	Utah	Salt Lake	19
7	2,003	8	Utah	Salt Lake	3
8	2,003	11	Utah	Salt Lake	1
9	2,003	12	Utah	Salt Lake	3

2013

```
select extract(year from recorded_date)
rec_year,
extract(month from recorded_date) rec_month,
state_name, county_name, count(category)
from "2013_aqi_data"
where state_name = 'Utah'
and county_name = 'Salt Lake'
and category like 'Unhealthy%'
group by extract(year from recorded_date),
extract(month from recorded_date), state_name,
county_name
```

select extract(year from recorded_c

Enter a SQL expression to filter results (u

	rec_year	rec_month	state_name	county_name	count
1	2,013	1	Utah	Salt Lake	17
2	2,013	2	Utah	Salt Lake	5
3	2,013	3	Utah	Salt Lake	1
4	2,013	6	Utah	Salt Lake	4
5	2,013	7	Utah	Salt Lake	9
6	2,013	11	Utah	Salt Lake	2
7	2,013	12	Utah	Salt Lake	12

2023

```
select extract(year from recorded_date)
rec_year,
extract(month from recorded_date) rec_month,
state_name, county_name, count(category)
from "2023_AQI_Data"
where state_name = 'Utah'
and county_name = 'Salt Lake'
and category like 'Unhealthy%'
group by extract(year from recorded_date),
extract(month from recorded_date), state_name,
county_name
```

select extract(year from recorded_c

Enter a SQL expression to filter results (use Ct

	rec_year	rec_month	state_name	county_name	count
1	2,023	2	Utah	Salt Lake	3
2	2,023	5	Utah	Salt Lake	1
3	2,023	6	Utah	Salt Lake	1
4	2,023	7	Utah	Salt Lake	4
5	2,023	8	Utah	Salt Lake	7

In Salt Lake county, in 2003, July had the most unhealthy days. But in 2013, it was December and January that had the most unhealthy days. In 2023, August had the most unhealthy days. The variation could be due to inversion in winter months and forest smoke fires during summer months.

Summary:

Dear Governor and Leaders of legislature, I am pleased to present my report on the air quality after conducting an analysis on the air quality index. This analysis shows that significant improvement were made with the number of unhealthy days in Utah County with zero unhealthy days in 2023. This is a remarkable improvement over the 30 unhealthy days in 2003 and the 41 unhealthy days in 2013.

The progress, however, in improving the unhealthy days in Salt Lake County is a bit more mixed. In 2003 and 2023, a few of the summer months had the most or one of the greatest number of unhealthy days. For example, July 2003 had the most unhealthy days at 19 days. August 2013 had 9 unhealthy days which was also the third highest number of unhealthiest days for a month in 2013. Finally, August 2023 had 7 unhealthy days which was the highest number of unhealthy days for a month in 2023. Unfortunately, this means that summer months are still the most unhealthiest months in the year. It should be noted that some of the contributing factors for summer days being the most unhealthiest may be out of control like wildfires that create smoke from California. Efforts can however still be made to combat this, like finding ways to reduce how much electricity Utahns consume in the summer for AC's which can help reduce pollution, since electricity generation creates more pollution.

Additionally, Salt Lake County has made great progress in reducing how many unhealthy days are in each month. For instance, in 2003, every month had at least one unhealthy day, while 2013 had 9 months with at least one unhealthy day. Finally, 2023 had 5 months with at least one unhealthy day which is really great progress.

Overall, the air quality seems to be improving, which indicates that current policies seem to be working in both of Utah's most populous counties, Salt Lake County and Utah County. Some policy changes may have to be implemented for the summer months to reduce the number of unhealthy days. Apart from that, the state seems to be doing a good job.

Thank you for taking the time to read this analysis.

Best Regards,

Varun Selvam