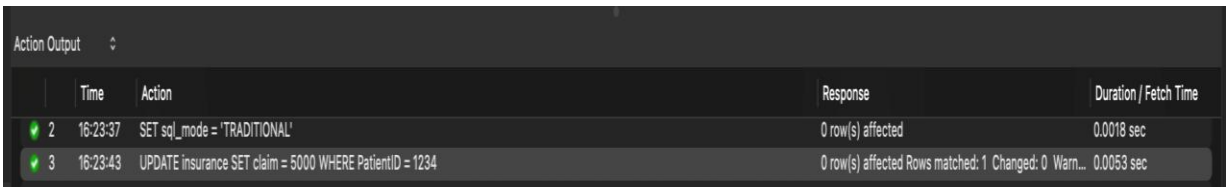
SQL Case Study on Insurance Data Set

```
USE nikhil;

SET sql_mode = 'TRADITIONAL';


-- 1. Write a SQL query to update the claim amount for the patient with PatientID = 1234 to 5000.


UPDATE insurance

SET claim = 5000

WHERE PatientID = 1234;
```




```
-- 2. Write a SQL query to delete all records for patients who are smokers and have no children.


DELETE

FROM insurance

WHERE smoker = 'Yes' AND children = 0;


16:30:15	DELETE  FROM insurance WHERE smoker = 'Yes' AND children = 0	0 row(s) affected	0.0031 sec
```

-- 3. Who are the top 5 patients who claimed the highest insurance amounts?

```
SELECT *, DENSE_RANK () OVER(ORDER BY claim DESC)
FROM insurance
LIMIT 5;
```

	index	PatientID	age	gender	bmi	bloodpressu...	diabetic	children	smoker	region	claim	DENSE_RANK () OVER(ORDER BY claim DESC)
	1337	1338	30	male	34.5	91	Yes	3	Yes	northwest	60021.4	1
	1336	1337	59	female	38.1	120	No	1	Yes	northeast	58571.07	2
	1333	1334	44	male	36.4	127	No	1	Yes	southwest	51194.56	3
	1332	1333	26	male	37	120	No	2	Yes	southeast	49577.66	4
	1331	1332	18	male	41.1	104	No	1	Yes	southeast	48970.25	5

-- 4. What is the average insurance claimed by patients based on the number of children they have?

```
SELECT children, ROUND(avg_claim, 2)
FROM (SELECT *, AVG(claim) OVER(PARTITION BY children) avg_claim, ROW_NUMBER()
OVER(PARTITION BY children) row_num FROM insurance) t
WHERE t.row_num = 1;
```

	children	ROUND(avg_claim, 2)
	0	7584.96
	1	12630.53
	2	15073.56
	3	15355.32
	4	13850.66
	5	8786.04

-- 5. What is the highest and lowest claimed amount by patients in each region?

```
SELECT *
FROM (SELECT *, MAX(claim) OVER(PARTITION BY region) max_claim,
                MIN(claim) OVER(PARTITION BY region) min_claim,
                ROW_NUMBER() OVER(PARTITION BY region) row_num
                FROM insurance) t
WHERE t.row_num =1;
```

index	PatientID	age	gender	bmi	bloodpressu...	diabetic	children	smoker	region	claim	max_claim	min_claim	row_num
15	16	32	male	30.4	86	Yes	0	No		1256.3	1256.3	1252.41	1
932	933	47	male	32.1	91	No	0	No	northeast	13555	58571.07	1694.8	1
207	208	48	female	40.2	82	No	0	No	northwest	3201.25	60021.4	1136.4	1
0	1	39	male	23.2	91	Yes	0	No	southeast	1121.87	49577.66	1121.87	1
260	261	28	male	26.8	92	No	3	No	southwest	3906.13	51194.56	1261.44	1

-- 6. What is the percentage of smokers in each age group?

```
SELECT age, COUNT(CASE WHEN smoker = 'yes' THEN 1 END) * 100.0 / COUNT(*) AS
smoker_percentage
FROM insurance
GROUP BY age
ORDER BY age;
```

age	smoker_percenta...
	0.00000
18	6.25000
19	18.51852
20	21.73913
21	17.64706
22	29.41176
23	12.50000
24	0.00000
25	10.71429
26	20.00000
27	11.11111
28	14.70588
29	13.88889
30	13.51351
31	5.55556
32	15.38462
33	6.25000
34	2.85714
35	6.06061
36	13.88889
37	15.38462
38	17.24138
39	24.00000

-- 7. What is the difference between the claimed amount of each patient and the first claimed amount of that patient?

```
SELECT *, claim - FIRST_VALUE(claim) OVER() difference_claim
FROM insurance;
```

index	PatientID	age	gender	bmi	bloodpressu...	diabetic	children	smoker	region	claim	difference_claim
0	1	39	male	23.2	91	Yes	0	No	southeast	1121.87	0
1	2	24	male	30.1	87	No	0	No	southeast	1131.51	9.6400000000001
2	3		male	33.3	82	Yes	0	No	southeast	1135.94	14.070000000000164
3	4		male	33.7	80	No	0	No	northwest	1136.4	14.5300000000002
4	5		male	34.1	100	No	0	No	northwest	1137.01	15.1400000000001
5	6		male	34.4	96	Yes	0	No	northwest	1137.47	15.600000000000136
6	7		male	37.3	86	Yes	0	No	northwest	1141.45	19.580000000000155
7	8	19	male	41.1	100	No	0	No	northwest	1146.8	24.930000000000064
8	9	20	male	43	86	No	0	No	northwest	1149.4	27.5300000000002
9	10	30	male	53.1	97	No	0	No	northwest	1163.46	41.590000000000146
10	11	36	male	19.8	88	Yes	0	No	northwest	1241.57	119.70000000000005

-- 8. For each patient, calculate the difference between their claimed amount and the average claimed amount of patients with the same number of children.

```
SELECT *, claim - AVG(claim) OVER(PARTITION BY children) avg_diff
FROM insurance;
```

index	PatientID	age	gender	bmi	bloodpressu...	diabetic	children	smoker	region	claim	avg_diff
0	1	39	male	23.2	91	Yes	0	No	southeast	1121.87	-6463.091366594359
1	2	24	male	30.1	87	No	0	No	southeast	1131.51	-6453.451366594359
2	3		male	33.3	82	Yes	0	No	southeast	1135.94	-6449.021366594359
3	4		male	33.7	80	No	0	No	northwest	1136.4	-6448.561366594358
4	5		male	34.1	100	No	0	No	northwest	1137.01	-6447.951366594359
5	6		male	34.4	96	Yes	0	No	northwest	1137.47	-6447.491366594359
6	7		male	37.3	86	Yes	0	No	northwest	1141.45	-6443.511366594359
7	8	19	male	41.1	100	No	0	No	northwest	1146.8	-6438.161366594359
8	9	20	male	43	86	No	0	No	northwest	1149.4	-6435.561366594358
9	10	30	male	53.1	97	No	0	No	northwest	1163.46	-6421.501366594359
10	11	36	male	19.8	88	Yes	0	No	northwest	1241.57	-6343.391366594359
11	12	37	male	20.3	90	Yes	0	No	northwest	1242.26	-6342.701366594359
12	13	19	male	20.7	81	No	0	No	northwest	1242.82	-6342.141366594359
13	14	32	male	27.6	100	No	0	No		1252.41	-6332.551366594359
14	15	40	male	28.7	81	Yes	0	No		1253.94	-6331.021366594359
15	16	32	male	30.4	86	Yes	0	No		1256.3	-6328.661366594359
16	17	35	male	34.1	90	No	0	No	southwest	1261.44	-6323.521366594359
17	18	41	male	34.4	84	No	0	No	southwest	1261.86	-6323.101366594359
18	19	49	male	35.4	97	Yes	0	No	southwest	1263.25	-6321.711366594359
19	20	48	male	33.3	91	Yes	0	No	southeast	1391.53	-6193.431366594359
20	21	45	male	23.2	85	Yes	0	No	southeast	1515.34	-6069.621366594359
21	22	34	male	31.1	96	No	0	No	southwest	1526.31	-6058.65136659435...
22	23	18	male	35.5	100	Yes	0	No	southeast	1532.47	-6052.491366594359

-- 9. Show the patient with the highest BMI in each region and their respective rank.

SELECT * FROM (SELECT *, ROW_NUMBER() OVER(PARTITION BY region ORDER BY bmi DESC) AS row_num, RANK() OVER(ORDER BY bmi DESC) AS ranking

FROM insurance) t

WHERE t.row_num = 1;

index	PatientID	age	gender	bmi	bloodpressu...	diabetic	children	smoker	region	claim	row_num	ranking
9	10	30	male	53.1	97	No	0	No	northwest	1163.46	1	1
1299	1300	50	male	52.6	110	No	1	Yes	southeast	44501.4	1	2
675	676	49	female	48.1	81	Yes	2	No	northeast	9432.93	1	5
1306	1307	43	female	47.6	112	Yes	2	Yes	southwest	46113.51	1	7
15	16	32	male	30.4	86	Yes	0	No		1256.3	1	612

-- 10. Calculate the difference between the claimed amount of each patient and the claimed amount of the patient who has the highest BMI in their region.

SELECT *, claim - FIRST_VALUE(claim) OVER(PARTITION BY region ORDER BY bmi DESC)

FROM insurance;

index	PatientID	age	gender	bmi	bloodpressu...	diabetic	children	smoker	region	claim	claim - FIRST_VALUE(claim) OVER(PARTITIO...
15	16	32	male	30.4	86	Yes	0	No		1256.3	0
14	15	40	male	28.7	81	Yes	0	No		1253.94	-2.3599999999999
13	14	32	male	27.6	100	No	0	No		1252.41	-3.8899999999998727
675	676	49	female	48.1	81	Yes	2	No	northeast	9432.93	0
681	682	27	female	44.7	98	No	0	No	northeast	9541.7	108.77000000000044
1267	1268	43	female	42.8	105	Yes	1	Yes	northeast	40904.2	31471.269999999997
1290	1291	47	male	41.9	140	Yes	3	Yes	northeast	43753.34	34320.409999999996
1023	1024	51	female	41.3	98	No	0	No	northeast	17878.9	8445.970000000001
817	818	35	male	40.9	80	Yes	0	No	northeast	11566.3	2133.369999999999
707	708	27	female	40.7	100	No	0	No	northeast	9875.68	442.75
1326	1327	26	male	40.6	113	Yes	3	Yes	northeast	48549.18	39116.25
1077	1078	19	male	40.3	110	No	0	No	northeast	20709.02	11276.09
692	693	44	female	40	96	No	3	No	northeast	9704.67	271.7399999999998
175	176	59	female	39.8	81	Yes	0	No	northeast	2755.02	-6677.91
1056	1057	21	male	39.7	107	Yes	4	No	northeast	19496.72	10063.79
599	600	34	male	39.4	94	Yes	1	No	northeast	8342.91	-1090.0200000000004
980	981	36	female	39.3	82	No	0	No	northeast	14901.52	5468.59
895	896	49	male	39.1	102	Yes	0	No	northeast	12890.06	3457.129999999999
804	805	36	female	38.4	98	Yes	2	No	northeast	11396.9	1963.9699999999993
568	569	58	female	38.3	87	Yes	0	No	northeast	7935.29	-1497.6400000000003
968	969	37	male	38.2	97	Yes	0	No	northeast	14410.93	4978
986	987	55	female	38.1	91	No	2	No	northeast	15230.32	5797.389999999999
1137	1138	58	female	38.1	94	No	2	No	northeast	24915.05	15482.119999999999

-- 11. For each patient, calculate the difference in claim amount between the patient and the patient with the highest claim amount among patients with the same bmi and smoker status, within the same region. Return the result in descending order difference.

```
SELECT *, (MAX(claim) OVER(PARTITION BY region, bmi, smoker) - claim) diff
FROM insurance
ORDER BY diff DESC;
```

index	PatientID	age	gender	bmi	bloodpressu...	diabetic	children	smoker	region	claim	diff
2	3		male	33.3	82	Yes	0	No	southeast	1135.94	35444.34
19	20	48	male	33.3	91	Yes	0	No	southeast	1391.53	35188.75
228	229	40	male	34.8	81	No	3	No	southwest	3443.06	33467.55
239	240	31	female	34.8	88	No	1	No	southwest	3578	33332.61
282	283	48	female	33.3	100	Yes	1	No	southeast	4151.03	32429.25
377	378	57	female	34.8	96	No	1	No	southwest	5246.05	31664.56
260	261	28	male	26.8	92	No	3	No	southwest	3906.13	31253.999999999996
406	407	43	female	33.3	94	Yes	1	No	southeast	5594.85	30985.43
474	475	55	female	34.8	80	Yes	2	No	southwest	6571.54	30339.07
594	595	33	female	33.3	93	Yes	0	No	southeast	8283.68	28296.6
210	211	59	female	28.6	82	No	0	No	northeast	3213.62	27046.38
87	88	37	male	25.8	86	No	0	No	northwest	1972.95	26950.19
743	744	42	male	33.3	101	Yes	3	No	southeast	10560.49	26019.79
152	153	18	male	27.6	94	Yes	0	No	northwest	2523.17	25817.019999999997
29	30	35	female	31.4	93	No	0	No	southeast	1622.19	25723.850000000002
78	79	42	female	31.5	100	No	0	No	southeast	1877.93	25123.05
120	121	37	female	31.9	89	No	0	No	southeast	2205.98	25116.75
110	111	27	female	27.1	88	Yes	0	No	southwest	2154.36	23986
0	1	39	male	23.2	91	Yes	0	No	southeast	1121.87	23959.9
381	382	44	female	25.8	90	No	0	No	northwest	5266.37	23656.77
20	21	45	male	23.2	85	Yes	0	No	southeast	1515.34	23566.43
101	102	33	male	27.4	100	No	0	No	northeast	2104.11	23552.47
191	192	28	female	24.1	80	No	0	No	southwest	2974.13	23262.45

-- 12. For each patient, find the maximum BMI value among their next three records (ordered by age).

```
SELECT *, MAX(bmi) OVER(ORDER BY age ROWS BETWEEN 1 FOLLOWING AND 3 FOLLOWING)
FROM insurance;
```

index	PatientID	age	gender	bmi	bloodpressu...	diabetic	children	smoker	region	claim	MAX(bmi) OVER(ORDER BY age ROWS BET...
2	3		male	33.3	82	Yes	0	No	southeast	1135.94	34.4
3	4		male	33.7	80	No	0	No	northwest	1136.4	37.3
4	5		male	34.1	100	No	0	No	northwest	1137.01	37.3
5	6		male	34.4	96	Yes	0	No	northwest	1137.47	37.3
6	7		male	37.3	86	Yes	0	No	northwest	1141.45	30.8
152	153	18	male	27.6	94	Yes	0	No	northwest	2523.17	30.8
723	724	18	male	30.8	94	No	3	No	northeast	10141.14	32.3
326	327	18	male	30.8	97	Yes	0	No	southwest	4646.76	36
564	565	18	male	26.6	95	No	1	No	southeast	7742.11	36
632	633	18	male	32.3	98	Yes	1	No	northwest	8765.25	36
517	518	18	male	36	93	No	2	No	southeast	7160.33	32
259	260	18	male	30.9	92	No	2	No	northwest	3877.3	32
890	891	18	male	29.6	81	Yes	0	No	northeast	12731	32
580	581	18	male	32	81	Yes	2	No	northwest	8116.27	29.7
41	42	18	male	27.8	93	No	0	No	northwest	1635.73	41.1
1185	1186	18	male	29.7	103	No	2	No	northeast	32108.66	41.1
244	245	18	male	25.5	99	Yes	0	No	northeast	3645.09	43.7
1331	1332	18	male	41.1	104	No	1	Yes	southeast	48970.25	43.7
462	463	18	male	29.8	97	No	2	No	southeast	6406.41	43.7
818	819	18	male	43.7	101	No	1	No	southwest	11576.13	35.5
22	23	18	male	35.5	100	Yes	0	No	southeast	1532.47	42.9
659	660	19	male	29.6	94	No	5	No	northeast	9222.4	42.9
1016	1017	19	male	23	99	Yes	2	Yes	northwest	17361.77	42.9

-- 13. For each patient, find the rolling average of the last 2 claims.

SELECT *, AVG(claim) OVER(ROWS BETWEEN 2 PRECEDING AND 1 PRECEDING)

FROM insurance;

index	PatientID	age	gender	bmi	bloodpressu...	diabetic	children	smoker	region	claim	AVG(claim) OVER(ROWS BETWEEN 2 PRECE...
0	1	39	male	23.2	91	Yes	0	No	southeast	1121.87	NULL
1	2	24	male	30.1	87	No	0	No	southeast	1131.51	1121.87
2	3		male	33.3	82	Yes	0	No	southeast	1135.94	1126.69
3	4		male	33.7	80	No	0	No	northwest	1136.4	1133.725
4	5		male	34.1	100	No	0	No	northwest	1137.01	1136.17
5	6		male	34.4	96	Yes	0	No	northwest	1137.47	1136.705
6	7		male	37.3	86	Yes	0	No	northwest	1141.45	1137.24
7	8	19	male	41.1	100	No	0	No	northwest	1146.8	1139.46
8	9	20	male	43	86	No	0	No	northwest	1149.4	1144.125
9	10	30	male	53.1	97	No	0	No	northwest	1163.46	1148.1
10	11	36	male	19.8	88	Yes	0	No	northwest	1241.57	1156.43
11	12	37	male	20.3	90	Yes	0	No	northwest	1242.26	1202.5149999999999
12	13	19	male	20.7	81	No	0	No	northwest	1242.82	1241.915
13	14	32	male	27.6	100	No	0	No		1252.41	1242.54
14	15	40	male	28.7	81	Yes	0	No		1253.94	1247.615
15	16	32	male	30.4	86	Yes	0	No		1256.3	1253.1750000000002
16	17	35	male	34.1	90	No	0	No	southwest	1261.44	1255.12
17	18	41	male	34.4	84	No	0	No	southwest	1261.86	1258.87
18	19	49	male	35.4	97	Yes	0	No	southwest	1263.25	1261.65
19	20	48	male	33.3	91	Yes	0	No	southeast	1391.53	1262.5549999999998
20	21	45	male	23.2	85	Yes	0	No	southeast	1515.34	1327.3899999999999
21	22	34	male	31.1	96	No	0	No	southwest	1526.31	1453.435
22	23	18	male	35.5	100	Yes	0	No	southeast	1532.47	1520.8249999999998

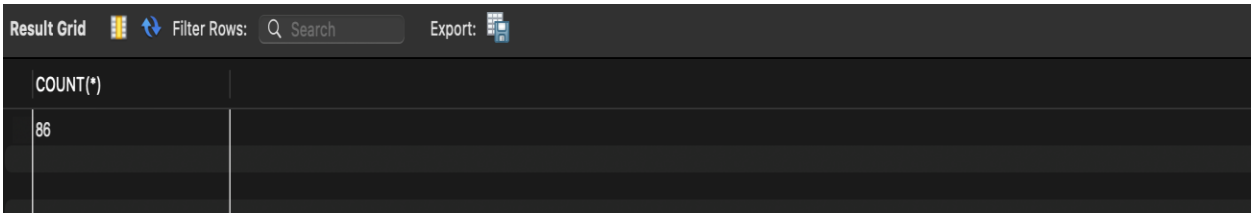
-- 14. Find the first claimed insurance value for male and female patients, within each region order the data by patient age in ascending order, and only include patients who are non-diabetic and have a bmi value between 25 and 30.

```
WITH t1 AS  (
            SELECT *
            FROM insurance
            WHERE diabetic = 'No' AND (bmi BETWEEN 25 AND 30))

SELECT region,gender,first_claim FROM (SELECT *,
FIRST_VALUE(claim) OVER(PARTITION BY region,gender ORDER BY age) AS first_claim,
ROW_NUMBER() OVER(PARTITION BY region,gender ORDER BY age) AS row_num
FROM t1) t
WHERE t.row_num = 1;
```

region	gender	first_claim
	male	1252.41
northeast	female	10106.13
northeast	male	32108.66
northwest	female	4189.11
northwest	male	1635.73
southeast	female	8219.2
southeast	male	7742.11
southwest	female	9861.03
southwest	male	25309.49

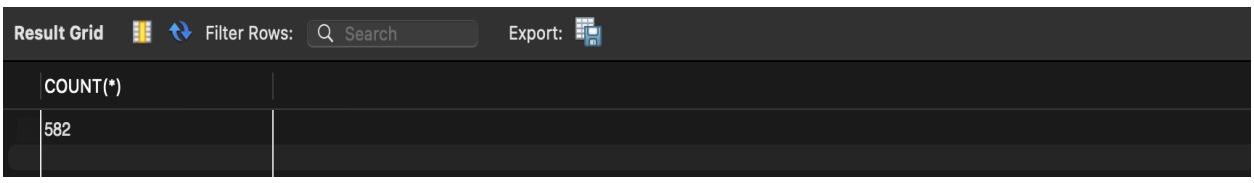
-- 15. How many patients have claimed more than the average claim amount for patients who are smokers and have at least one child and belong to the southeast region?

```
SELECT COUNT(*)
FROM insurance
WHERE claim > (SELECT AVG(claim) avg_claim
                FROM insurance
                WHERE smoker = 'yes' AND children >= 1 AND region = 'southeast');
```



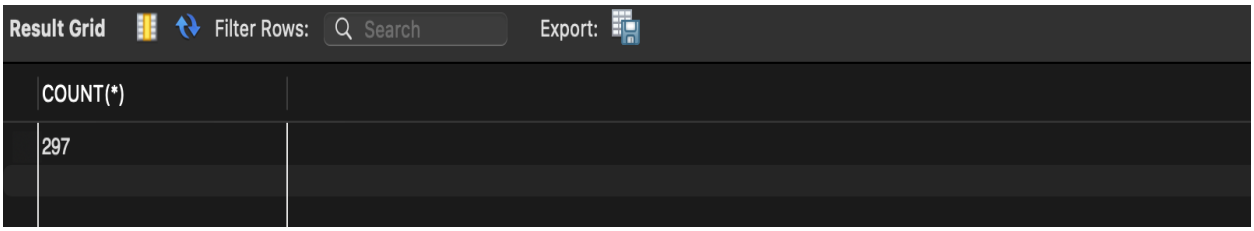
-- 16. How many patients have claimed more than the average claim amount for patients who are not smokers and have a BMI greater than the average BMI for patients who have at least one child?

```
SELECT COUNT(*)
FROM insurance
WHERE claim > (SELECT AVG(claim)
                FROM insurance
                WHERE smoker = 'No' AND bmi > (SELECT AVG(bmi)
                                                FROM insurance
                                                WHERE children >= 1));
```



-- 17. How many patients have claimed more than the average claim amount for patients who have a BMI greater than the average BMI for patients who are diabetic, have at least one child, and are from the southwest region?

```
SELECT COUNT(*)
FROM insurance
WHERE claim > (SELECT AVG(claim)
                    FROM insurance
                    WHERE bmi > (SELECT AVG(BMI)
                                FROM insurance
                                WHERE diabetic = 'yes' AND children >= 1 AND
                                region = 'southwest'));
```



-- 18. What is the difference in the average claim amount between patients who are smokers and patients who are non-smokers and have the same BMI and number of children?

```
SELECT AVG(t1.claim - t2.claim) difference_claim
FROM (SELECT *
            FROM insurance
            WHERE smoker = 'Yes') t1
```

```
JOIN (SELECT *
                FROM insurance
                WHERE smoker = 'No') t2
ON t1.BMI = t2.BMI AND t1.children = t2.children;
```