Health Insurance SQL Documentation

1) Count the number of Policyholders

```
10    # b) Count the number of policyholders
11 •  SELECT COUNT(*) AS total_policyholders FROM insurance;
12
```

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

total_policyholders
1338

2) Total claim amounts by region

```
13    # Total claim amounts by region
14 •  SELECT region, SUM(charges) AS total_charges
15    FROM insurance
16    GROUP BY region;
17
```

Result Grid | Filter Rows: | Export: | Wrap

region	total_charges
southwest	4012754.647620001
southeast	5363689.763290002
northwest	4035711.9965399993
northeast	4343668.583308999

3)Average claim amount based on smoking status

```
18    #Average claim amount based on smoking status
19 •  SELECT smoker, AVG(charges) AS avg_charges
20    FROM insurance
21    GROUP BY smoker;
22
```

Result Grid | Filter Rows: | Export: | Wrap C

smoker	avg_charges
yes	32050.23183153285
no	8434.268297856199

4)Identify the region with the highest average claim amount

```
23      # Identify the region with the highest average claim amount
24 •    SELECT region, AVG(charges) AS avg_charges
25      FROM insurance
26      GROUP BY region
27      ORDER BY avg_charges DESC
28      LIMIT 1;
29
30
```

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

region	avg_charges
southeast	14735.411437609895

5)Analyzr the impact of BMI on claim amounts

```
30      #Analyze the impact of BMI on claim amounts
31 •    SELECT
32          CASE
33              WHEN bmi < 18.5 THEN 'Underweight'
34              WHEN bmi BETWEEN 18.5 AND 24.9 THEN 'Normal weight'
35              WHEN bmi BETWEEN 25 AND 29.9 THEN 'Overweight'
36              ELSE 'Obese'
37          END AS bmi_category,
38          AVG(charges) AS avg_charges
39      FROM insurance
40      GROUP BY bmi_category;
```

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

bmi_category	avg_charges
Overweight	10993.994037132627
Obese	15479.549772628647
Normal weight	10379.499732162163
Underweight	8852.200585000002

6) Average claim amount in each region by smoking status

```
42      #Region-Specific Analysis
43      # Average claim amount in each region by smoking status
44 •    SELECT region, smoker, AVG(charges) AS avg_charges
45      FROM insurance
46      GROUP BY region, smoker;
47
```

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

region	smoker	avg_charges
southwest	yes	32269.0634936207
southeast	no	8032.216308937727
northwest	no	8556.463715205993
northeast	no	9165.531671708171
southeast	yes	34844.99682362637
southwest	no	8019.284513071162
northeast	yes	29673.536472835818
northwest	yes	30192.00318241379

7)Region – specific highest claim amounts

```
48      #Region-specific highest claim amounts
49 •    SELECT region, MAX(charges) AS max_charges
50      FROM insurance
51      GROUP BY region;
52
53
```

Result Grid | Filter Rows: | Export: | Wrap

region	max_charges
southwest	52590.82939
southeast	63770.42801
northwest	60021.39897
northeast	58571.07448

8)Find key attributes influencing higher claims

```
54      #v a) Find key attributes influencing higher claims
55 •    SELECT
56          age,
57          sex,
58          bmi,
59          children,
60          smoker,
61          charges
62      FROM insurance
63      WHERE charges > (SELECT AVG(charges) FROM insurance)
64      ORDER BY charges DESC;
```

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

age	sex	bmi	children	smoker	charges
49	female	27.1	1	no	26140.3603
20	female	24.42	0	yes	26125.67477
56	male	26.695	1	yes	26109.32905
21	female	32.68	2	no	26018.95052
52	male	26.4	3	no	25992.82104
59	female	23.655	0	yes	25678.77845
50	female	27.36	0	no	25656.57526
54	male	25.46	1	no	25517.11363
54	male	25.1	3	yes	25382.297

8)Region – based filtering for further modelling

```
66      # Region-based filtering for further modeling
67 •    SELECT *
68      FROM insurance
69      WHERE region = 'southeast';
70
```

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

age	sex	bmi	children	smoker	region	charges
18	male	33.77	1	no	southeast	1725.5523
28	male	33	3	no	southeast	4449.462
31	female	25.74	0	no	southeast	3756.6216
46	female	33.44	1	no	southeast	8240.5896
62	female	26.29	0	yes	southeast	27808.7251
56	female	39.82	0	no	southeast	11090.7178
27	male	42.13	0	yes	southeast	39611.7577
18	male	34.1	0	no	southeast	1137.011
59	female	27.72	3	no	southeast	14001.1338
31	female	36.63	2	no	southeast	4949.7587
41	male	21.78	1	no	southeast	6272.4772
37	female	30.8	2	no	southeast	6313.759
60	female	24.53	0	no	southeast	12629.8967
36	male	35.2	1	yes	southeast	38709.176

9) Total claim amounts by region and smoking status

```
72      # Total claim amounts by region and smoking status
73 •    SELECT region, smoker, SUM(charges) AS total_charges
74      FROM insurance
75      GROUP BY region, smoker;
76
```

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

region	smoker	total_charges
southwest	yes	1871605.6826300006
southeast	no	2192795.0523399995
northwest	no	2284575.81196
northeast	no	2355541.639629
southeast	yes	3170894.7109499993
southwest	no	2141148.9649900002
northeast	yes	1988126.9436799998
northwest	yes	1751136.18458

10)Age distribution of policyholders

```
77    #Age distribution of policyholders
78 •  SELECT age, COUNT(*) AS policyholder_count
79    FROM insurance
80    GROUP BY age
81    ORDER BY age;
82
```

Result Grid | Filter Rows: | Export: | Wrap

age	policyholder_count
18	69
19	68
20	29
21	28
22	28
23	28
24	28
25	28
26	28
27	28
28	28
29	27
30	27
31	27
32	26