Data

The following schema diagram shows the tables available.

Task 1

Column Name	Criteria
id	Discrete. The unique identifier of the support ticket. </br>Missing values are not possible due to the database structure.
customer_id	Discrete. The unique identifier of the customer. </br>Missing values should be replaced with 0.
category	Nominal. The gategory of the support request, can be one of Feedback, Billing Enquiry, Bug, Installation Problem, Other. </br>Missing values should be replaced with Other.
status	Nominal. The current status of the support ticket, one of Open, In Progress or Resolved. </br>Missing values should be replaced with 'Resolved'.
creation_date	Discrete. The date the ticket was created. Can be any date in 2023. </br>Missing values should be replaced with 2023-01-01.
response_time	Discrete. The number of days taken to respond to the support ticket. </br>Missing values should be replaced with 0.
resolution_time	Continuos. The number of hours taken to resolve the support ticket, rounded to 2 decimal places. </br>Missing values should be replaced with 0.

In [240]:

```
SELECT
    id,
    COALESCE(customer_id, 0) AS customer_id,
    COALESCE(
        CASE
            WHEN category IN ('Feedback', 'Billing Enquiry', 'Bug', 'Installation
Problem', 'Other') THEN category
            ELSE 'Other'
        END,
        'Other'
    ) AS category,
    COALESCE(
        CASE
            WHEN status IN ('Open', 'In Progress', 'Resolved') THEN status
            ELSE 'Resolved'
        END,
        'Resolved'
    ) AS status,
    COALESCE(
        CASE
            WHEN creation_date BETWEEN '2023-01-01' AND '2023-12-31' THEN
creation_date
            ELSE '2023-01-01'::DATE
        END,
        '2023-01-01'::DATE
    ) AS creation_date,
    COALESCE(response_time, 0) AS response_time,
    ROUND(
        COALESCE(
```

```
            CASE
                WHEN resolution_time ~ '^[0-9]+(\.[0-9]+)?$' THEN resolution_time::
NUMERIC
                WHEN resolution_time ~ '^[0-9]+(\.[0-9]+)? hours$' THEN
REPLACE(resolution_time, ' hours', '')::NUMERIC
                ELSE NULL
            END,
            0
        ),
        2
    ) AS resolution_time
FROM
    support
WHERE
    id IS NOT NULL;
```

Out[240]:

	id	customer_id	category	status	creation_date	response_time	resolution_time
0	1	1062	Installation Problem	In Progress	2023-01-26 00:00:00+00:00	6	0.00
1	2	892	Other	Open	2023-06-18 00:00:00+00:00	3	0.00
2	3	433	Feedback	Open	2023-08-17 00:00:00+00:00	1	0.00
3	6	764	Other	Open	2023-01-16 00:00:00+00:00	3	0.00
4	7	1144	Other	Open	2023-06-01 00:00:00+00:00	2	0.00
...	...	...	...	...	...	...	...
1982	2995	1091	Other	Resolved	2023-10-01 00:00:00+00:00	3	3.93
1983	2996	1024	Bug	In Progress	2023-07-31 00:00:00+00:00	1	0.00
1984	2997	1105	Installation Problem	In Progress	2023-02-06 00:00:00+00:00	5	0.00
1985	2998	1608	Bug	In Progress	2023-06-10 00:00:00+00:00	3	0.00
1986	3000	1087	Installation Problem	Open	2023-02-15 00:00:00+00:00	8	0.00

1987 rows × 7 columns

Task 2

It is suspected that the response time to tickets is a big factor in unhappiness.

Calculate the minimum and maximum response time for each category of support ticket.

Expected output should include the columns `category`, `min_response` and `max_response`.

Values should be rounded to two decimal places where appropriate.

In [241]:

```
-- Write your query for task 2 in this cell
```

```
SELECT
 category,  --select category
 round(min(response_time),2) as min_response, --minimum of response time
 round(max(response_time),2) as max_response  --max of response time
FROM support
GROUP BY category    --group by unique
```

Out[241]:

	category	min_response	max_response
0	Other	1.0	5.0
1	Bug	1.0	13.0
2	Feedback	1.0	2.0
3	Billing enquiry	2.0	8.0
4	Installation Problem	5.0	17.0

Task 3

The support team want to know more about the `rating` provided by customers who reported `Bugs` or `Installation Problem`s.

Expected query to return the `rating` from the survey, the `customer_id`, `category` and `response_time` of the support ticket, for the customers & categories of interest.

In [242]:

```
-- Write your query for task 3 in this cell
SELECT
 su.rating,
    s.customer_id,
    s.category,
    s.response_time
FROM
    survey AS su
INNER JOIN
    support as s
ON
    su.customer_id = s.customer_id
WHERE
    s.category IN ('Bug', 'Installation Problem');
```

Out[242]:

	rating	customer_id	category	response_time
0	6	621	Installation Problem	7
1	5	1703	Installation Problem	6
2	5	766	Installation Problem	7
3	5	1824	Bug	3
4	4	931	Installation Problem	9
...	...	...	...	...
118	4	1371	Installation Problem	6
119	4	1434	Installation Problem	6

	rating	customer_id	category	response_time
120	3	1228	Bug	4
121	3	926	Installation Problem	6
122	6	1227	Installation Problem	5

123 rows × 4 columns