PROJECT: SQL ASSOCIATE PRACTICAL EXAM

datalab

Practical Exam: Hotel Operations

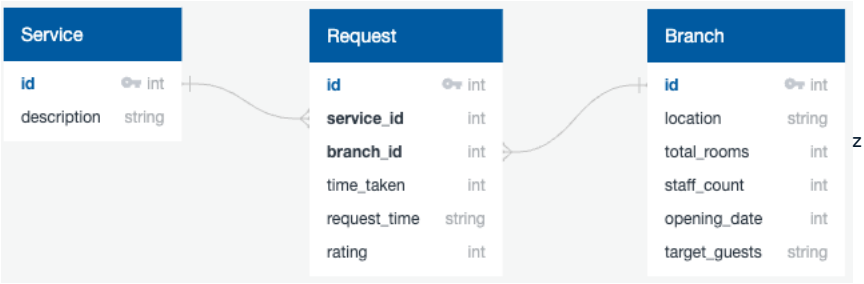
LuxurStay Hotels is a major, international chain of hotels. They offer hotels for both business and leisure travellers in major cities across the world. The chain prides themselves on the level of customer service that they offer.

However, the management has been receiving complaints about slow room service in some hotel branches. As these complaints are impacting the customer satisfaction rates, it has become a serious issue. Recent data shows that customer satisfaction has dropped from the 4.5 rating that they expect.

You are working with the Head of Operations to identify possible causes and hotel branches with the worst problems.

Data

The following schema diagram shows the tables available. You have only been provided with data where customers provided a feedback rating.



z

Task 1

Before you can start any analysis, you need to confirm that the data is accurate and reflects what you expect to see.

It is known that there are some issues with the `branch` table, and the data team have provided the following data description.

Write a query to return data matching this description, including identifying and cleaning all invalid values. You must match all column names and description criteria. Your output should be a DataFrame named 'clean_branch_data'.

Column Name	Criteria
id	Nominal. The unique identifier of the hotel. Missing values are not possible due to the database structure.
location	Nominal. The location of the particular hotel. One of four possible values, 'EMEA', 'NA', 'LATAM' and 'APAC'. Missing values should be replaced with “Unknown”.
total_rooms	Discrete. The total number of rooms in the hotel. Must be a positive integer between 1 and 400. Missing values should be replaced with the default number of rooms, 100.
staff_count	Discrete. The number of staff employeed in the hotel service department. Missing values should be replaced with the total_rooms multiplied by 1.5.
opening_date	Discrete. The year in which the hotel opened. This can be any value between 2000 and 2023. Missing values should be replaced with 2023.
target_guests	Nominal. The primary type of guest that is expected to use the hotel. Can be one of 'Leisure' or 'Business'. Missing values should be replaced with 'Leisure'.

Unknown integration DataFrame as `clean_branch_data`

```
SELECT
    id,
    COALESCE(location, 'Unknown') AS location,
    COALESCE(NULLIF(total_rooms, 0), 100) AS total_rooms,
    COALESCE(NULLIF(staff_count, 0), (total_rooms * 1.5)) AS staff_count,
    COALESCE(NULLIF(
        CASE WHEN opening_date = '-' THEN NULL ELSE CAST(opening_date AS INT) END
    , 0), 2023) AS opening_date,
    COALESCE(
        CASE WHEN target_guests = 'B.' THEN 'Business' ELSE target_guests END,
        'Leisure'
    ) AS target_guests
FROM branch;
```

ind...	id	locati...	total_rooms	staff_count	opening_date	target_guests
0	1	LATAM	168	178	2017	Business
1	2	APAC	154	82	2010	Leisure
2	3	APAC	212	467	2003	Leisure
3	4	APAC	230	387	2023	Business
4	5	APAC	292	293	2002	Business
5	6	NA	260	590	2022	Leisure
6	7	EMEA	259	442	2018	Business
7	8	NA	259	285	2023	Business
8	9	NA	157	274	2001	Business
9	10	EMEA	205	138	2013	Leisure
10	11	EMEA	191	255	2005	Business
11	12	NA	177	248	2012	Business
12	13	EMEA	126	255	2010	Leisure
13	14	EMEA	366	703	2000	Business
14	15	APAC	365	688	2002	Business
15	16	LATAM	228	274	2021	Business

Rows: 100 Expand

Task 2

The Head of Operations wants to know whether there is a difference in time taken to respond to a customer request in each hotel. They already know that different services take different lengths of time.

Calculate the average and maximum duration for each branch and service.

- Your output should be a DataFrame named 'average_time_service'
- It should include the columns `service_id`, `branch_id`, `avg_time_taken` and `max_time_taken`
- Values should be rounded to two decimal places where appropriate.

Unknown integration DataFrame as a

```
-- Write your query for task 2 in this cell
select service_id, branch_id, Round(avg(time_taken),2) as avg_time_taken, max(time_taken) as max_time_taken
from public.request
Group by service_id,branch_id;
```

	s...	b...	avg_tim...	max_ti...
0	2	46	13.09	16
1	4	99	9.13	13
2	1	8	2.56	10
3	2	13	13.53	17
4	1	46	2.08	4
5	3	15	6.73	7
6	2	35	13.17	16
7	1	1	2.44	12
8	3	13	6.8	8
9	1	57	2.29	5
10	1	41	2.27	8
11	2	32	13.23	19
12	4	66	9	10
13	1	23	2.46	9
14	3	22	7.15	9
15	3	57	7.81	13

Rows: 385 Expand

Task 3

The management team want to target improvements in `Meal` and `Laundry` service in Europe (`EMEA`) and Latin America (`LATAM`).

Write a query to return the `description` of the service, the `id` and `location` of the branch, the id of the request as `request_id` and the `rating` for the services and locations of interest to the management team.

Your output should be a DataFrame named 'target_hotels'.

Use the original branch table, not the output of task 1.

Unknown integration DataFrame as t

```
-- Write your query for task 3 in this cell
SELECT
description,
branch.id AS id,
location,
request.id AS request_id,
rating
FROM request
INNER JOIN service ON request.service_id = service.id
INNER JOIN branch ON request.branch_id = branch.id
WHERE (description IN ('Meal','Laundry'))
AND (location IN ('EMEA' , 'LATAM'))
```

	des...			r...	
0	Laundry	63	EMEA	3	4
1	Laundry	69	LATAM	6	5
2	Meal	44	EMEA	18	4
3	Laundry	57	LATAM	19	3
4	Meal	1	LATAM	21	4
5	Meal	26	LATAM	26	5
6	Laundry	34	EMEA	27	4
7	Laundry	60	LATAM	35	4
8	Meal	21	EMEA	37	4
9	Meal	1	LATAM	38	4
10	Meal	26	LATAM	41	5
11	Laundry	30	EMEA	44	5
12	Meal	21	EMEA	51	4
13	Laundry	69	LATAM	55	5
14	Meal	70	LATAM	63	4
15	Meal	23	EMEA	66	5

Rows: 5,047 Expand

Task 4

So that you can take a more detailed look at the lowest performing hotels, you want to get service and branch information where the average rating for the branch and service combination is lower than 4.5 - the target set by management.

- Your output should be a DataFrame named 'average_rating'
- It should return the `service_id` and `branch_id`, and the average rating (`avg_rating`)
- Values should be rounded to 2 decimal places where appropriate.

Unknown integration DataFrame as

```sql
-- Write your query for task 4 in this cell
select branch_id, service_id, round (avg(rating),2)  as avg_rating
from request
Group by branch_id, service_id
having avg(rating) < 4.5
order by avg_rating asc;
```

	b...	s...	a...
0	39	4	3
1	100	2	3
2	20	4	3
3	99	2	3.29
4	8	3	3.38
5	97	1	3.5
6	20	3	3.5
7	9	3	3.5
8	55	4	3.5
9	55	1	3.52
10	57	3	3.53
11	97	2	3.53
12	86	4	3.55
13	87	4	3.56
14	5	2	3.56
15	64	4	3.56

Rows: 215 Expand