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.

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. You must match all column names and description criteria.

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,
location,
COALESCE(total_rooms :: INTEGER,100) AS total_rooms,
staff_count :: DECIMAL,
REPLACE (opening_date :: TEXT,'-','2023') :: INTEGER AS opening_date,
REPLACE(REPLACE (target_guests :: TEXT, 'Busniess','Business'),'B.','Business') AS target_guests
FROM branch;
```

	id	location	total_rooms	staff_count	opening_date
0	1	LATAM	168	178	
1	2	APAC	154	82	
2	3	APAC	212	467	
3	4	APAC	230	387	
4	5	APAC	292	293	
5	6	NA	260	590	
6	7	EMEA	259	442	
7	8	NA	259	285	
8	9	NA	157	274	
9	10	EMEA	205	138	
10	11	EMEA	191	255	
11	12	NA	177	248	
12	13	EMEA	126	255	
13	14	EMEA	366	703	
14	15	APAC	365	688	

100 rows

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 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 `average_time_service`

```
-- 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 request GROUP BY service_id,
branch_id;
```

	service_id	branch_id	avg_time_taken
0	2	46	13.
1	4	99	9.
2	1	8	2.
3	2	13	13.
4	1	46	2.
5	3	15	6.
6	2	35	13.
7	1	1	2.
8	3	13	
9	1	57	2.
10	1	41	2.
11	2	32	13.
12	4	66	
13	1	23	2.
14	3	22	7.

385 rows

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.

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

Unknown integration DataFrame as `t`

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

	description	branch_id	location	request_id
0	Laundry	63	EMEA	
1	Laundry	69	LATAM	
2	Meal	44	EMEA	
3	Laundry	57	LATAM	
4	Meal	1	LATAM	
5	Meal	26	LATAM	
6	Laundry	34	EMEA	
7	Laundry	60	LATAM	
8	Meal	21	EMEA	
9	Meal	1	LATAM	
10	Meal	26	LATAM	
11	Laundry	30	EMEA	
12	Meal	21	EMEA	
13	Laundry	69	LATAM	
14	Meal	70	LATAM	

5,047 rows

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 query should return the `service_id` and `branch_id`, and the average rating (`avg_rating`), rounded to 2 decimal places.

Unknown integration DataFrame as a

```sql
-- Write your query for task 4 in this cell
SELECT service_id, branch_id, ROUND(AVG(rating), 2) AS avg_rating FROM request GROUP BY service_id, branch_id HAVING AVG(rating)<4.5;
```

	service_id	branch_id	avg_
0	2	46	
1	4	99	